SEMI-ANNUAL STATUS REPORT

Error Control Techniques for Sattellite and
Space Communications
NASA Grant Number NAG5-557

Principal Investigator:
Daniel J. Costello, Jr.

March 1989

Summary of Progress

During the period June 1, 1988 - December 31, 1988, progress was made in the following areas:

1) Performance Analysis of Bandwidth Efficient Trellis Codes on Channels with Phase Jitter

In our January 1988 semi-annual report, we included some preliminary results on the effect of phase jitter on trellis coded phase modulation. We are now in the process of preparing a paper for submission to the *IEEE Transactions on Communications* summarizing our results on phase jitter. A preliminary version of this paper is included as Appendix A of this report [1].

Figure 9 in Appendix A contains a good summary of our results for coded 8-PSK and 16-QAM modulation. When channel noise is the only cause of the phase jitter, i.e., the case of practical coherence, the SNR degradation due to phase jitter is less than with uncoded modulation. In other words, coding makes the system less sensitive to phase jitter. For example, with a phase noise standard deviation of 5°, coded 8-PSK suffers a degradation of ≈ 0.5 dB, whereas uncoded 4-PSK suffers a degradation of ≈ 1 dB. On the other hand, when oscillator noise is also a factor, the performance degradation can be much worse, as illustrated in Figure 11 of Appendix A.

Our work on the performance analysis of bandwidth efficient trellis codes on channels with phase jitter was performed by Dr. Christian Schlegel, a former Ph.D. student supported by the grant, who graduated in December 1988. Dr. Schlegel is now employed by ABB Corporate Research, Baden, Switzerland.

2) Performance Analysis of Bandwidth Efficient Trellis Codes on Channels Disturbed by Jamming and Impulse Noise

In our June 1988 annual report, we included some results on the performance analysis of bandwidth efficient trellis codes on fading channels. We have now extended these results to include channels disturbed by jamming and impulse noise. A paper containing these results was presented at the "Trois Journees sur le Codage" conference held in Toulon, France in November 1988. A copy of this paper is included as Appendix B of this report [2].

In this paper we analyze the performance of trellis coded modulation on channels with jamming and impulse noise. We assume a partial band jammer with an optimized duty cycle as a worst case impulse noise interferer. It is shown that the effective length, or minimum Hamming distance, of the code is the most important code parameter in determining performance. This is the same design parameter that resulted in the case of fading channels. However, the secondary design parameter differs in the two cases. Whereas for fading channels the minimum product distance was the secondary design parameter, for impulse noise channels the minimum free Euclidean distance fills this role. Some new trellis codes are constructed

for impulse noise channels using these new design criteria. In Figures 5-7 of Appendix B, it is shown that they perform better than codes designed for the AWGN channel.

Our work on the performance analysis of bandwidth efficient trellis codes on channels disturbed by jamming and impulse noise was performed by Dr. Christian Schlegel, a former Ph.D. student supported by the grant who graduated in December 1988. Dr. Schlegel is now employed by ABB Corporate Research, Baden, Switzerland.

3) A New Construction Algorithm for Trellis Codes

In our January 1988 semi-annual report, we included preliminary results on some hueristic construction algorithms for bandwidth efficient trellis codes with large constraint lengths. One of these algorithms has been developed further and has resulted in the construction of some good, long convolutional and trellis codes. A paper containing these results was presented at the "Trois Journees sur le Codage" conference held in Toulon, France in November 1988. A copy of this paper is included as Appendix C of this report [3].

The algorithm can be used to construct good convolutional or trellis codes with any constraint length up to about 30, any signal constellation, and any code rate. No exhaustive search is required, so the complexity of the algorithm depends only on the difficulty of computing the free distance as the constraint length increases. As shown in Tables 1 and 2 of Appendix C, the codes constructed are either optimum or nearly optimum in free distance. The construction of long constraint length convolutional and trellis codes is important for applications requiring extremely low decoded bit error rates, such as NASA deep space missions. It is anticipated that these long constraint length codes would be decoded using sequential decoding. The application of sequential decoding to bandwidth efficient trellis codes will be the subject of future research under this grant.

Our work on this construction algorithm for trellis codes was performed by Dr. Marc Rouanne, a former Ph.D. student supported by the grant who graduated in May 1988. Dr. Rouanne is now employed by Matra Communication, Paris, France.

4) Construction of Good Distance Profile Trellis Codes for Sequential Decoding

Another approach to constructing good trellis codes for use with sequential decoding is to optimize the distance profile rather than the free distance. Chevillat and Costello [4] showed that this approach worked well for binary convolutional codes. We have taken a similar approach to the construction of bandwidth efficient trellis codes for sequential decoding. A paper containing our results was presented at the Allerton Conference on Communication, Control, and Computing held in Monticello, Illinois in September 1988. A more complete version of the same paper is in the process of being prepared for submission to the *IEEE Transactions on Communications.* A preliminary version of this paper is included as Appendix D of this report [5].

Using a nested step-by-step code search algorithm, we constructed 2-, 4-, and 8-dimensional trellis codes using rate $k/k+1$ systematic binary convolutional codes with 8-PSK and 16-QAM

signal sets. The Fano sequential decoding algorithm was simulated to study the effects of the distance profile on the computational load of a sequential decoder. An examination of Figures 3, 7, and 8 in Appendix D illustrates that codes with a better distance profile can be decoded much faster using a sequential decoder. These results are important when selecting good codes for use with sequential decoding.

Our work on the construction of good distance profile codes for use with sequential decoding was performed by Mr. Sankar Malladi, a former M.S. student supported by the grant, who graduated in December 1988.

5) Comparison of Random Coding Bounds of Trellis Coded Modulation Schemes

In our January 1988 semi-annual report, we summarized our work on minimum free Euclidean distance lower bounds for trellis coded modulation. Since then we have obtained a related upper bound on the event error probability of trellis coded modulation. A paper containing this new upper bound along with a comparison to the lower bounds on free distance was presented at the European Signal Processing Conference held in Grenoble, France in September 1988. A copy of this paper is included as Appendix E of this report [6].

This new bound allows us to optimize a signal constellation for use with trellis coding by varying the signal points in the constellation until the best bound on event error probability is achieved. Figure 4, 5, and 6 in Appendix E show several examples of optimized signal constellations. These results are useful when one is attempting to find the signal constellation which will give the best performance with trellis coding.

Our work on finding the best signal constellations to use with trellis coding was performed by Dr. Christian Schlegel, a former Ph.D. student supported by the grant, who graduated in December 1988. Dr. Schlegel is now employed by ABB Corporate Research, Baden, Switzerland.

References

[1] C. Schlegel and D. J. Costello, Jr., "Trellis Coded Modulation on Channels with Phase Jitter", to be submitted to *IEEE Trans. Commun.*.

[2] C. Schlegel and D. J. Costello, Jr., "Bandwidth Efficient Coding on Channels Disturbed by Jamming and Impulse Noise", Proc. Colloque "Trois Journees sur le Codage", Toulon, France, November 1988.

[3] M. Rouanne and D. J. Costello, Jr., "A Construction Algorithm for Trellis Codes", Proc. Colloque "Trois Journees sur le Codage", Toulon, France, November 1988.

[4] P. R. Chevillat and D. J. Costello, Jr., "Distance and Computation in Sequential Decoding", *IEEE Trans. Commun.*, COM-24, pp. 440-447, April 1976.

[5] S. S. Malladi, D. J. Costello, Jr., and Hendrik C. Ferreira, "Construction of Good Distance Profile Trellis Codes for Sequential Decoding", to be submitted to *IEEE Trans. Commun.*.

[6] C. Schlegel, M. Rouanne, and D. J. Costello, Jr., "A Comparison of the Random Coding Bounds on the Free Euclidean Distance and the Event Error Probability of Trellis Coded Modulation Schemes", Signal Processing IV: Theories and Applications, pp. 1525-1528, Elsevier Science Publishers, 1988.

Appendix A

Trellis Coded Modulation on Channels with Phase Jitter

Trellis Coded Modulation on Channels with Phase Jitter [1]

Christian Schlegel

ABB Corporate Research
5405 Baden
Switzerland

Daniel J. Costello, Jr.

Department of Electrical and Computer Engineering
University of Notre Dame
Notre Dame, IN 46556

Submitted to the IEEE Transactions on Communications
March, 1989

abstract

Bandwidth efficient data transmission using trellis coded modulation (TCM) relies on coherent reception. Non-ideal carrier synchronization causes performance degradation. In this paper we apply the Chernoff bounding technique to obtain performance bounds for bandwidth efficient trellis codes on channels with phase jitter. It is shown that under the assumption of independent phase errors the performance degradation with respect to ideal synchronization is small. The independence of successive phase errors can be achieved in two ways: signal interleaving or the use of a decision-feedback predictor to whiten the phase errors. In the case of practical coherence, where channel noise is the only cause of phase jitter, a generalized distance is defined and generalized distance spectra are presented. The event error probability can be calculated for quasi-regular codes with an efficient algorithm . It is shown that this method can also be used to evaluate the influence of different causes of phase jitter. This is illustrated for the case of oscillator phase noise.

[1]This work was supported by NASA Grant NAG 5-557.

1 Introduction

In a coherent trellis coded digital communication system, two essential functions must be performed by the receiver before the received two-dimensional signal vectors, denoted by $\underline{y}_r$, can be passed on to the decoder. First, the carrier wave must be synchronized, i.e., the locally generated reference waveforms must be accurately phase locked to those produced at the transmitter. This is called *carrier recovery.* Second, symbol timing must be recovered. This is referred to as *timing recovery.* The timing recovery circuit extracts a sinusoid of frequency $1/T$ from the baseband signal, where T is the baseband signal duration. The sinusoid's phase determines the sampling instant and is important to proper system operation. Uncertainty in the frequency of the timing clock can usually be kept very small and timing recovery is less critical in terms of performance sensitivity than carrier recovery [1].

The carrier recovery circuit produces the local carrier signal $\cos(\omega_0 t + \theta(t))$, where ω_0 is the carrier frequency and $\theta(t)$ is the phase error process. This local carrier signal should be sufficiently accurate so that the phase error is small in both its static (mean) value and in its fluctuations, which are called *phase jitter.* There exist several methods to recover the carrier frequency [2]. One method consists of extracting the carrier frequency with a *phase-locked loop* (PLL). Another, decision-directed recovery, operates as follows. Due to disturbances, the receiver generated signal points $\underline{y}_r$ will not lie at the discrete points of the signal constellation, but rather in small disc like regions around those signal points. The angular deviation of the received signal vectors $\underline{y}_r$ from the points of the signal constellation generates sequences of control signals which are low-pass-filtered and applied to a voltage controlled oscillator, which then generates the local carrier frequency and phase.

Decision directed carrier recovery is usually used for TCM due to its better tracking performance [3]. At the sampling instants the phase error process $\theta(t)$ leads to phase offset angles θ_r (Figure 1). The θ_r are approximately Gaussian distributed and strongly correlated due to the low pass filtering in the carrier recovery circuit [2,4]. The strong correlation has led to the assumption of a constant phase offset over the length of an error event in TCM. Hagenauer and Sundberg [5] show that such a constant phase offset results in serious degradation in system performance.

In this paper we assume independent phase offset angles θ_r and show that the performance degradation is then relatively small. The independence of successive phase error angles can be achieved by interleaving, which is used to randomize the sequence of chan-

nel signals. We show in Appendix A that a linear phase predictior may also be used to whiten the residual phase errors. This preserves the advantages of decision directed phase sznchronization which have to be given up with interleaving. We will henceforth assume that the phase errors are independent.

The phase errors can also have a static component, but it can usually be made quite small by the recovery circuit. It is much harder to minimize the fluctuations of the carrier phases, which result not only from channel noise but also from other physical phenomena.

Figure 1: Effects of the r-th phase offset angle θ_r on the received signal constellation.

2 Performance Bounds

In this section we present a general method for deriving error performance bounds for coded systems used on memoryless channels[2]. We will apply these results to TCM communication systems whose structure is shown in Figure 2. A TCM communication system consists of a trellis encoder, possibly a signal interleaver, the transmission channel, possibly a signal deinterleaver, and a trellis decoder. A rate $R = k/n$ trellis code is generated by a binary convolutional encoder followed by a mapper. The convolutional encoder is a finite state automaton with 2^ν possible states, where ν is the memory order of the encoder. At each time interval r, the encoder accepts $\tilde{k}$ binary input bits $(u_r^{\tilde{k}}, u_r^{\tilde{k}-1}, \ldots, u_r^1)$ and makes a transition from its state S_r at time r to one of $2^{\tilde{k}}$ possible successor states S_{r+1}. The $\tilde{n} = n - (k - \tilde{k})$ output bits of the convolutional encoder and the $k - \tilde{k}$ uncoded information

[2]This method can be extended to also include finite state channels.

bits $(u_r^k, \ldots, u_r^{\tilde{k}+1})$ form one of 2^n binary n-tuples $v_r = (v_r^n, v_r^{n-1}, \ldots, v_r^1)$, which is translated by the mapper into one of $A = 2^n$ channel signals from a signal set $\mathcal{A} = \{\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_A\}$.

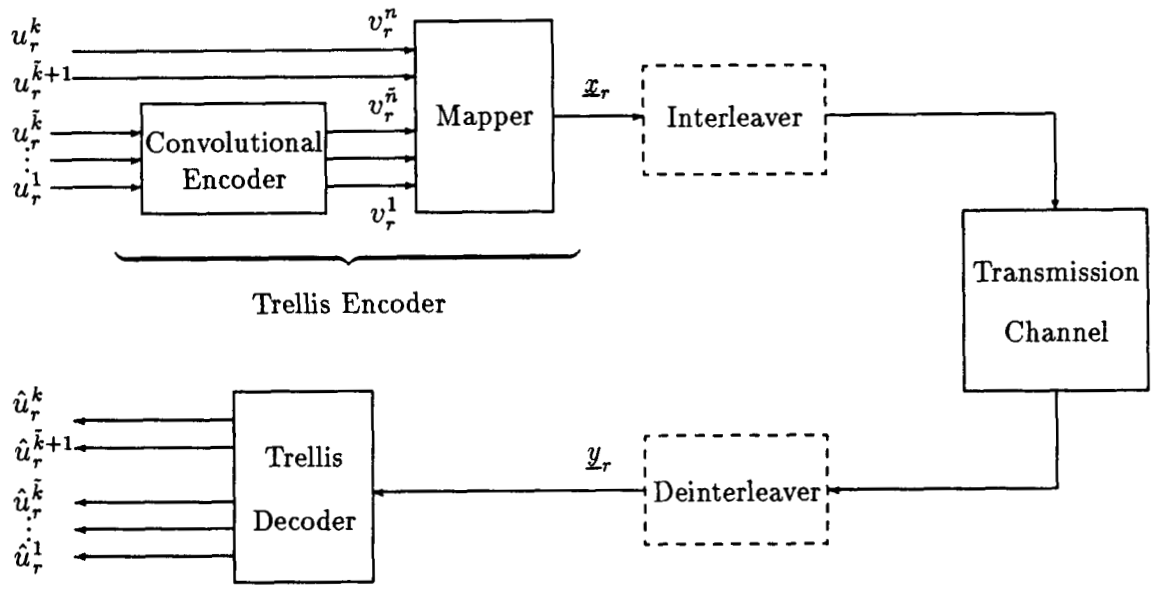


Figure 2: Trellis coded communication system.

The uncoded information bits do not affect the state of the convolutional encoder and cause $2^{k-\tilde{k}}$ parallel transitions between the encoder states S_r and S_{r+1}. Since the coherent DSB-AM system transmits two dimensions in one analog waveform signal, it is sensible to design the trellis encoder for 2-dimensional signal sets. A rate $R = k/n$ trellis code transmits k bits/channel signal, where the channel signal set contains $A = 2^n$ signals. If such a TCM communication system replaces an uncoded system that uses a signal set with $A' = 2^k$ signals, the overall transmission rate is preserved, and we call such a TCM system bandwidth efficient.

The particular transmission channel discussed in this paper is the AWGN-channel suffering from phase jitter. The linear phase offset predictor or the interleaver/deinterleaver insure that the phase offset angles θ_j in the received sequence are independent. They are decoded by a maximum likelihood sequence estimator (usually using the Viterbi algorithm). The Viterbi algorithm finds the signal sequence that most closely corresponds to the sequence of received signals. It achieves this by calculating the decoding metric $m(\mathbf{x}, \mathbf{y})$

between $\mathbf{x}$ and $\mathbf{y}$, where $\mathbf{x} = (\underline{x}_1, \ldots, \underline{x}_l)$ is a possible sequence of 2-dimensional transmitted signals and $\mathbf{y} = (\underline{y}_1, \ldots, \underline{y}_l)$ is the received sequence. $m(\mathbf{x}, \mathbf{y})$ is some non-negative function of $\mathbf{x}$ given $\mathbf{y}$, which is inversely related to the conditional probability that $\mathbf{x}$ was transmitted if $\mathbf{y}$ was received. The decoder will then choose the message sequence $\mathbf{x}$ for which this metric is minimized. It makes an error if it decodes a sequence $\mathbf{x}'$, given that the correct sequence, i.e., the transmitted sequence, was $\mathbf{x}$. This will happen if $m(\mathbf{x}', \mathbf{y}) \leq m(\mathbf{x}, \mathbf{y})$.

We assume at each step of the derivation that the receiver has no information about the phase error, since if it had such information, the received 2-dimensional signals $\underline{y}_r$ could easily be rotated by the erroneous phase angle into the correct position. We first calculate the two code word error probability conditioned on the phase error θ, and then average over the density function of θ. Denote the two code word error probability when $\mathbf{x}$ is sent and $\mathbf{x}'$ is received, given the phase errors $\theta = (\theta_1, \theta_2, \ldots, \theta_l)$ at each symbol interval, by $P(\mathbf{x} \rightarrow \mathbf{x}' \mid \theta)$. For the benefit of a more streamlined mathematical notation, we let $\underline{x}_r$ be a complex number of which the real part is the in-phase component $\hat{I}$ of the signal and the imaginary part is the quadrature component $\hat{Q}$. We further assume without loss of generality that the signals have unit average energy, i.e., $E_S = \frac{1}{A} \sum_p |\underline{a}_p|^2 = 1$. In any case, we may always normalize the signal constellation and the channel noise by E_S without changing any of the derived expressions.

The two code word error probability, i.e., the probability that $\mathbf{x}'$ is erroneously decoded if $\mathbf{x}$ is sent is given by

$$\begin{aligned} P(\mathbf{x} \rightarrow \mathbf{x}') &= \underset{\theta}{E}[P(\mathbf{x} \rightarrow \mathbf{x}' \mid \theta)] \\ P(\mathbf{x} \rightarrow \mathbf{x}') &= \underset{\theta}{E} \Pr\{m(\mathbf{x}', \mathbf{y}) - m(\mathbf{x}, \mathbf{y}) \leq 0 \mid \theta\}. \end{aligned} \tag{1}$$

We use the Chernoff bounding technique [7] to upper bound the above expression and obtain

$$\Pr\{m(\mathbf{x}', \mathbf{y}) - m(\mathbf{x}, \mathbf{y}) \leq 0 \mid \theta\} \leq \underset{\mathbf{y}|\mathbf{x}}{E} [\exp(-\lambda\{m(\mathbf{x}', \mathbf{y}) - m(\mathbf{x}, \mathbf{y})\}) \mid \theta], \tag{2}$$

where $\underset{\mathbf{y}|\mathbf{x}}{E}$ denotes conditional expectation and λ is a non-negative real valued parameter over which we minimize the right hand side of (2) to obtain the tightest possible exponential bound, i.e.,

$$P(\mathbf{x} \rightarrow \mathbf{x}') \leq \min_{\lambda} \underset{\theta}{E} \underset{\mathbf{y}|\mathbf{x}}{E} [\exp(-\lambda\{m(\mathbf{x}', \mathbf{y}) - m(\mathbf{x}, \mathbf{y})\})]$$

$$= \min_{\lambda} C(\mathbf{x}, \mathbf{x}', \lambda), \tag{3}$$

where $C(\mathbf{x}, \mathbf{x}', \lambda)$ is called the Chernoff bound between the signal sequences $\mathbf{x}$ and $\mathbf{x}'$.

Restricting attention to decoders using the additive maximum likelihood metric for the AWGN-channel, i.e.,

$$m(\mathbf{x}, \mathbf{y}) = \sum_{r=1}^{l} |\underline{y}_r - \underline{x}_r|^2 \tag{4}$$

we may rewrite (3) as

$$\begin{aligned} P(\mathbf{x} \to \mathbf{x}') &\leq \min_{\lambda} C(\mathbf{x}, \mathbf{x}', \lambda) \\ &= \min_{\lambda} \underset{\theta}{E} \underset{\mathbf{y}}{E} \left[\prod_{r=1}^{l} \exp(\lambda(|\underline{x}_r|^2 - |\underline{x}'_r|^2)) \exp(-2\lambda Re\{\underline{y}_r(\underline{x}_r - \underline{x}'_r)^*\}) \right] \\ &= \min_{\lambda} \prod_{r=1}^{l} C(\underline{x}_r, \underline{x}'_r, \lambda), \end{aligned} \tag{5}$$

where $C(\underline{x}_r, \underline{x}'_r, \lambda)$ is called the *Chernoff factor* of the signals $\underline{x}_r$ and $\underline{x}'_r$ and * denotes complex conjugation. Due to the incorrectly received carrier phase, the received signal $\underline{y}_r$ is given by $\underline{y}_r = e^{i\theta_r}\underline{x}_r + \underline{n}_r$, where $\underline{n}_r$ denotes the usual 2-dimensional Gaussian noise vector expressed as a complex number, and i denotes the imaginary unit $\sqrt{-1}$. After further manipulations we obtain

$$C(\underline{x}_r, \underline{x}'_r, \lambda) = \underset{\theta_r}{E} \left[\exp(\lambda(|\underline{x}_r|^2 - \underline{x}'_r|^2) + \lambda^2 N_0 |\underline{x}_r - \underline{x}'_r|^2 - 2\lambda Re\{e^{i\theta_r}\underline{x}_r(\underline{x}_r - \underline{x}'_r)^*\}) \right]. \tag{6}$$

In the next steps we will average (6) over the phase offset angles θ_r in the phase jitter case. The phase offset angles θ_r are statistically independent and identically distributed (i.i.d.). We then proceed to obtain for the Chernoff factor of the signals $\underline{x}_r$ and $\underline{x}'_r$:

$$C(\underline{x}_r, \underline{x}'_r, \lambda) = \exp(\lambda(|\underline{x}_r|^2 - |\underline{x}'_r|^2) + \lambda^2 N_0 |\underline{x}_r - \underline{x}'_r|^2) \underset{\theta}{E} \left[\exp(-2\lambda Re\{e^{i\theta}\underline{x}_r(\underline{x}_r - \underline{x}'_r)^*\}) \right]. \tag{7}$$

Due to the stationarity and independence of the phase error process θ, we have dropped the dummy subscript r in the averaging over θ above.

As discussed in the introduction, the phase offset angles θ_r are assumed to be Gaussian distributed, i.e.,

$$p(\theta) \approx \frac{1}{\sqrt{2\pi\sigma_\theta^2}} e^{-\theta^2/2\sigma_\theta^2}, \tag{8}$$

where σ_θ^2 is the phase offset angle variance.

The analysis method presented here is not limited to phase noise induced by channel noise. Any combination of sources of synchronization errors can be included as long as the net interference can be modeled as a stochastic process which has Gaussian statistics.

We now evaluate

$$\underset{\theta}{E}\Big[\exp(-2\lambda Re\{e^{i\theta}\underline{x}_r(\underline{x}_r - \underline{x}_r')^*\})\Big] = \underset{\theta}{E}[\exp(-2\lambda Re\{(\cos\theta + i\sin\theta)\underline{x}_r(\underline{x}_r - \underline{x}_r')^*\})]. \tag{9}$$

For phase error angles θ that are not too large, we may approximate the trigonometric functions by

$$\cos\theta \approx 1 - \frac{\theta^2}{2} \tag{10}$$

$$\sin\theta \approx \theta. \tag{11}$$

Using these expressions in the above evaluation of the expectation over θ, we obtain

$$\begin{aligned}
&\underset{\theta}{E}\Big[\exp(-2\lambda Re\{e^{i\theta}\underline{x}_r(\underline{x}_r - \underline{x}_r')^*\})\Big] = \\
&\quad \underset{\theta}{E}\left[\exp(-2\lambda Re\{(1 - \frac{\theta^2}{2} + i\theta)\underline{x}_r(\underline{x}_r - \underline{x}_r')^*\})\right] \\
&= \exp(-2\lambda Re\{\underline{x}_r(\underline{x}_r - \underline{x}_r')^*\})\underset{\theta}{E}\left[\exp(-2\lambda Re\{(-\frac{\theta^2}{2} + i\theta)\underline{x}_r(\underline{x}_r - \underline{x}_r')^*\})\right],
\end{aligned} \tag{12}$$

and $\underset{\theta}{E}[\cdot]$ is evaluated as

$$\underset{\theta}{E}[\cdot] \leq \frac{\exp\left(\frac{\lambda^2 Re\{i\underline{x}_r(\underline{x}_r - \underline{x}_r')^*\}^2}{1/2\sigma_\theta^2 - \lambda Re\{\underline{x}_r(\underline{x}_r - \underline{x}_r')^*\}}\right)}{\sqrt{1 - 2\sigma_\theta^2\lambda Re\{\underline{x}_r(\underline{x}_r - \underline{x}_r')^*\}}}. \tag{13}$$

Rewriting the signals $\underline{x}_r$ and $\underline{x}_r'$ in polar form, i.e.,

$$\begin{aligned} \underline{x}_r &= |\underline{x}_r| e^{i\phi_r} \\ \underline{x}'_r &= |\underline{x}'_r| e^{i\phi_{r'}}, \end{aligned} \tag{14}$$

we obtain

$$\begin{aligned} Re\{\mathrm{i}\underline{x}_r(\underline{x}_r - \underline{x}'_r)^*\} &= -Im\{\underline{x}_r(\underline{x}_r - \underline{x}'_r)^*\} \\ &= -Im\{\underline{x}_r\underline{x}_r^*\} + Im\{\underline{x}_r\underline{x}_r^{'*}\} \\ &= |\underline{x}_r||\underline{x}'_r| \sin(\phi_{rr'}), \end{aligned} \tag{15}$$

and analogously

$$Re\{\underline{x}_r(\underline{x}_r - \underline{x}'_r)^*\} = |\underline{x}_r|^2 - |\underline{x}_r||\underline{x}'_r| \cos(\phi_{rr'}), \tag{16}$$

where $\phi_{rr'} = \phi_r - \phi'_r$ is the angle between the two signals $\underline{x}_r$ and $\underline{x}'_r$. With these simplifications we obtain for (13)

$$\underset{\theta}{E}[\cdot] \leq \frac{1}{\sqrt{1 - 2\lambda\sigma_\theta^2\eta_r}} \exp\left(\frac{2\lambda^2\sigma_\theta^2\kappa_r^2}{1 - 2\lambda\sigma_\theta^2\eta_r}\right), \tag{17}$$

where $\kappa_r = |\underline{x}_r||\underline{x}'_r| \sin(\phi_{rr'})$ and $\eta_r = (|\underline{x}_r|^2 - |\underline{x}_r||\underline{x}'_r| \cos(\phi_{rr'}))$. Using (12) and (17) in (7), we obtain for the Chernoff factors

$$C(\underline{x}_r, \underline{x}'_r, \lambda) = \exp(-\lambda(1 - \lambda N_0)|\underline{x}_r - \underline{x}'_r|^2) \frac{\exp\left(\frac{2\lambda^2\sigma_\theta^2\kappa_r^2}{1-2\lambda\sigma_\theta^2\eta_r}\right)}{\sqrt{1 - 2\lambda\sigma_\theta^2\eta_r}}. \tag{18}$$

The first exponential expression is the Chernoff factor for the AWGN-channel. The second term depends on the phase error θ and reflects the *degradation* which a real system suffers from phase jitter with respect to an ideally synchronized system.

The Chernoff factors are important because they not only streamline the expression for the two code word error probability but also apply to the transfer function bound for trellis codes introduced later and to the cutoff rate calculations. In particular, it can be shown [8] that R_0, the channel cutoff-rate in bits/transmitted signal, is given by

$$R_0(p) = -\log_2 \min_{\lambda} \sum_{m=1}^{A} \sum_{p=1}^{A} p(\underline{a}_m) p(\underline{a}_p) C(\underline{a}_m, \underline{a}_p, \lambda), \tag{19}$$

where $p(\underline{a}_j)$ is the probability of choosing the signal $\underline{a}_j \in \mathcal{A}$. Note that R_0 is dependent on the particular metric $m(\underline{y}_r, \underline{x}_r)$ that is used by the decoder. If the decoder uses the maximum likelihood (ML)-metric for a memoryless channel, i.e.,

$$\begin{aligned} m(\mathbf{y},\mathbf{x}) &= -\log(\Pr(\mathbf{y}|\mathbf{x})) = -\log \prod_{r=1}^{l} \Pr(\underline{y}_r|\underline{x}_r) \\ &= \sum_{r=1}^{l} \Big(-\log(\Pr(\underline{y}_r|\underline{x}_r))\Big) = \sum_{r=1}^{l} m(\underline{x}_r, \underline{y}_r), \end{aligned} \tag{20}$$

(19) becomes the channel cutoff-rate for the optimum receiver, which is the usual definition of R_0 [7]. We will denote the value of λ which maximizes the cutoff-rate (19) by λ_{R_0}. In this case, the Chernoff factors will be written as $C(\underline{a}_m, \underline{a}_p) \triangleq C(\underline{a}_m, \underline{a}_p, \lambda_{R_0})$.

Evaluating (19) numerically, we observed that the minimizing value of λ, λ_{R_0}, is very close to the minimizing value of λ for the pure AWGN channel for $\sigma_\theta^2/N_0 \lessapprox 0.25$, i.e., $\lambda_{R_0} \approx 1/2N_0$. We therefore weaken the bound (17) and the Chernoff factors (18) minimally by setting $\lambda = 1/2N_0 \approx \lambda_{R_0}$ and the Chernoff factors become independent of λ and are given by

$$C(\underline{x}_r, \underline{x}'_r) \approx \exp\left(-\frac{1}{4N_0}|\underline{x}_r - \underline{x}'_r|^2\right) \frac{\exp\left(\frac{\sigma_\theta^2 \kappa^2}{2N_0^2 - 2N_0\sigma_\theta^2 \eta}\right)}{\sqrt{1 - \frac{\sigma_\theta^2 \eta}{N_0}}}. \tag{21}$$

The Chernoff factor in the general form of (21) is a function of the noise power spectral density N_0. The functional dependency of σ_θ has not yet been specified at this point.

3 Practical Coherence

Following the terminology of [4] we call a system *practically coherent* if the phase noise of the carrier recovery loop contains only a channel noise induced component. Assuming that a first-order phase-locked loop is used to extract the carrier phase from the discrete carrier component ,we may use the results of [4, p. 35] for the phase jitter angle variance, i.e.,

$$\sigma_\theta^2 = \frac{N_0 B_L}{E_S}, \tag{22}$$

where B_L is the *loop-noise bandwidth* of the phase locked loop. The same expression for the phase jitter angle variance is obtained for suppressed-carrier tracking using an A-th order phase locked loop as discussed in [4, chapter 2] and [9].

We now introduce the *phase-noise-to-channel-noise ratio* $\rho = \sigma_\theta^2 E_S/N_0$. Note that for the two PLL examples cited above, $\rho = B_L$, the loop-noise bandwidth. For PLL's, $\frac{E_S}{N_0 \rho}$ is also called the loop signal-to-noise ratio[3]. If a nonlinear model of the phase locked loop is employed, it can be shown [4, chapter 4] that (8) is reasonably accurate for $\sigma_\theta^2 < 0.25$. The Chernoff factor (21) can then be written as

$$C(\underline{x}_r, \underline{x}'_r) = \exp\left(-\frac{1}{4N_0}\left(|\underline{x}_r - \underline{x}'_r|^2 - \frac{2\rho\kappa^2}{1-\rho\eta}\right)\right)\frac{1}{\sqrt{1-\rho\eta}}. \tag{23}$$

The value of ρ relates the phase jitter of a particular receiver to the channel signal-to-noise ratio. A smaller value of ρ gives a better receiver, since the phase noise variance σ_θ^2 becomes smaller.

Although the bit error probability P_b is the quantity of ultimate interest, a closely related and more readily determined quantity, the *event error probability* P_e, will be used to characterize the performance of trellis codes.

If $\mathbf{x}$ and $\mathbf{x}'$ are two symbol sequences corresponding to two paths through the trellis which are distinct for l branches starting at node j, and the decoder chooses the encoded sequence $\mathbf{x}'$ over the correct sequence $\mathbf{x}$, this is called an error event of length l starting at node j. An error event starts where the two paths diverge and ends where the two paths remerge. A union bound on P_e for a trellis code may be obtained by summing the probabilities of the error events of all possible lengths given a particular correct sequence $\mathbf{x}$ and averaging this quantity over all possible correct sequences $\mathbf{x}$.

With each incorrect path we may associate a sequence of incorrect trellis states S'_r, while the sequence of correct states is denoted S_r. Any error event of length l can then be described by l state pairs, $(S_0, S'_0), \ldots, (S_l, S'_l)$, with $S_0 = S'_0$, $S_l = S'_l$, and $S_r \neq S'_r$ for $0 < r < l$, i.e., the incorrect path must not touch the correct path during the error event. Associated with these paths are the two symbol sequences $\mathbf{x} = (\underline{x}_0, \underline{x}_1, \ldots, \underline{x}_l)$ and $\mathbf{x}' = (\underline{x}'_0, \underline{x}'_1, \ldots, \underline{x}'_l)$, where $\underline{x}_r, \underline{x}'_r \in \mathcal{A}$. The probability of an error event may be upper bounded using the Chernoff factors (with $\lambda = \lambda_{R_0}$) between the individual signals of the two code sequences which the paths generate. We may therefore write

$$\Pr[(S_0, \ldots, S_l) \to (S'_0, \ldots, S'_l)] \leq C(\mathbf{x}, \mathbf{x}') = \prod_{r=1}^{l} C(\underline{x}_r, \underline{x}'_r). \tag{24}$$

[3]See [4, p. 33 ff.] for graphs of the pdf of θ for several loop signal-to-noise ratios.

Using (23) for the Chernoff factor we obtain

$$\Pr[(S_0,\ldots,S_l)\rightarrow(S',\ldots,S'_l)]\leq \tag{25}$$

$$\prod_{r=1}^{l}\exp\left(-\frac{1}{4N_0}\left(|\underline{x}_r-\underline{x}'_r|^2-\frac{2\rho\kappa_r^2}{1-\rho\eta}\right)\right)\frac{1}{\sqrt{1-\rho\eta}}$$

$$=\exp\left(-\frac{1}{4N_0}\sum_{r=1}^{l}\left(|\underline{x}_r-\underline{x}'_r|^2-\frac{2\rho\kappa_r^2}{1-\rho\eta_r}\right)\right)\prod_{r=1}^{l}\frac{1}{\sqrt{1-\rho\eta_r}}$$

$$=\mathcal{K}_r\exp\left(-\frac{1}{4N_0}d(\mathbf{x},\mathbf{x}')\right), \tag{26}$$

where $\mathcal{K}_r=\prod_{r=1}^{l}\frac{1}{\sqrt{1-\rho\eta_r}}$ is independent of the noise power spectral density N_0.

It becomes evident that the probability of choosing $\mathbf{x}'$ over the correct sequence $\mathbf{x}$ is determined by a *generalized distance* measure $d(\mathbf{x},\mathbf{x}')$ between the two sequences $\mathbf{x}$ and $\mathbf{x}'$, where

$$d(\mathbf{x},\mathbf{x}')=\sum_{r=1}^{l}d(\underline{x}_r,\underline{x}'_r)=\sum_{r=1}^{l}\left(|\underline{x}_r-\underline{x}'_r|^2-\frac{2\rho\kappa_r^2}{1-\rho\eta_r}\right), \tag{27}$$

which after substituting for κ_r and η_r becomes

$$d(\mathbf{x},\mathbf{x}')=\sum_{r=1}^{l}\left(|\underline{x}_r-\underline{x}'_r|^2-\frac{2\rho|\underline{x}_r|^2|\underline{x}'_r|^2\sin^2(\phi_{rr'})}{1-\rho(|\underline{x}_r|^2-|\underline{x}_r||\underline{x}'_r|\cos(\phi_{rr'}))}\right). \tag{28}$$

Note that the distance measure (28) is an additive distance measure and reduces to the Euclidean distance for negligible phase noise, i.e., for $\rho\rightarrow 0$.

For MPSK, the generalized distance in (28) can be reduced further to obtain

$$d(\mathbf{x},\mathbf{x}')=\sum_{r=1}^{l}\left(2(1-\cos(\phi_{rr'}))-\frac{2\rho\sin^2(\phi_{rr'})}{1-\cos(\phi_{rr'})}\right)$$

$$=\sum_{r=1}^{l}\left(2(1-\cos(\phi_{rr'}))\frac{1-2\rho}{1-\cos(\phi_{rr'})}\right). \tag{29}$$

From (29) it is evident that the generalized distance is never negative for $\rho<0.5$. It is also evident that close signals, i.e., small values of $\phi_{rr'}$, suffer more from phase jitter than signals which are further apart.

Tables 1 and 2 show the normalized minimum free generalized distances of the best known 8-PSK and 16-QAM trellis codes, which were found by Ungerboeck [3]. The entries list the normalized minimum free generalized distance $d(\mathbf{x}, \mathbf{x}')$ of the codes for various values of ρ. Note that the entry for $\rho = 0$ is the ordinary minimum free Euclidean distance d_{free}.

i.d.	ν	$H(D)^0$	$H(D)^1$	$H(D)^2$	$\rho = 0$	$\rho = 0.01$	$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.25$
u2	2	5	2	-	4	4	4	4	2.9827
u4	4	23	04	16	5.17	5.1111	4.8596	4.5211	3.2987
u5	5	45	16	34	5.75	5.6869	5.3946	5.0039	3.6148
u6	6	103	030	066	6.34	6.2686	5.9296	5.4866	3.9308
u7	7	277	054	122	6.58	6.5152	6.2193	5.8161	4.2468
u8	8	435	072	130	7.52	7.4141	6.9997	6.4522	4.5629
u9	9	1007	164	260	7.52	7.4141	6.9997	6.4522	4.5629
u10	10	2003	164	770	7.52	7.4141	6.9997	6.4522	4.5629

Table 1: Ungerboeck's 8-PSK codes on an AWGN-channel with phase jitter.

i.d.	ν	$H(D)^0$	$H(D)^1$	$H(D)^2$	$\rho = 0$	$\rho = 0.01$	$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.25$
uq2	= 2	5	2	-	1.6	1.5709	1.4468	1.2727	0.5714
uq3	3	11	02	04	2	1.9671	1.8339	1.6599	1.0159
uq4	4	23	04	16	2.4	2.3598	2.1959	1.9833	1.2041
uq5	5	41	06	10	2.4	2.3599	2.1967	1.9864	1.3099
uq6	6	101	016	064	2.8	2.7526	2.5588	2.3067	1.4597
uq7	7	203	014	042	3.2	3.1436	2.9209	2.6269	1.5416
uq8	8	401	056	304	3.2	3.1453	2.9209	2.6269	1.5416
uq9	9	1001	0346	0510	3.2	3.1870	3.1304	2.9566	1.7744

Table 2: Ungerboeck's 16-QAM codes on an AWGN-channel with phase jitter.

4 Performance of TCM on Channels with Phase Jitter

The performance of a code on an AWGN channel depends on its distance spectrum. The generalized distance measure introduced in the last section for channels with phase jitter changes this distance spectrum. Furthermore, for all but constant envelope constellations, the generalized branch distance $d(\underline{x}_r, \underline{x}'_r)$ is no longer reciprocal, i.e., $d(\underline{a}_m, \underline{a}_p) \neq d(\underline{a}_p, \underline{a}_m)$ for some $\underline{a}_m, \underline{a}_p \in \mathcal{A}$. The distance spectrum can be efficiently computed only for *quasi-regular* codes [10]. Quasi-regularity is defined as follows. Let S and S' be two states in a trellis code and let the *signal selector error vector* e be the binary sum of a signal selector

v from state S and a signal selector v' from S', i.e, $e = v \oplus v'$. The distance polynomial $P_{S,S',e}(x)$ is then defined as the set of distances between signals leaving S and S' whose signal selectors differ by e. A trellis code is quasi-regular if and only if any non-zero distance polynomial $P_{S,S',e}$ depends only on e but not on S and S'. In [10] it is shown that all $k/k+1$ trellis codes which are designed using Ungerboeck's set partitioning [3] are quasi-regular.

The next two lemmas establish that these codes are also quasi-regular with respect to the generalized distance measure introduced earlier.

Lemma 1 *8-PSK codes based on Ungerboeck's set partitioning are quasi-regular with respect to the generalized distance of (29).*

Proof: The generalized distance is reciprocal for constant envelope schemes, which can be seen by inspecting (29). Thus, only the values of the actual distances between points are altered by using the generalized distance instead of the Euclidean distance and the quasi-regularity of these 8-PSK codes for the Euclidean distance is a sufficient condition for quasi-regularity for the generalized distance. Q.E.D.

Lemma 2 *16-QAM codes based on Ungerboeck's set partitioning are quasi-regular with respect to the generalized distance of (28).*

Proof: Figure 3 shows the first level of the set partitioning of 16-QAM. The two sets are labeled B1 and B0 in accordance with Ungerboeck's notation. Codes based on set partitioning select signals either from set B1 or B0 at each step, regardless of the encoder state. Four possible combinations of the trellis states S and S' can occur. Furthermore, $e \in \{0000, 0010, 0100, 0110, 1000, 1010, 1100, 1110\}$ only if $(S, S') \in B1 \times B1$ or $(S, S') \in B0 \times B0$. From Figure 3 it can be seen that inverting the last bit in the signal selectors axially reflects the set $B1$ into $B0$ and vice versa. This reflection will not affect the magnitude $|\phi_{rr'}|$ of the angle of the signals with respect to each other nor the vector lengths $|\underline{x}_r|$ or $|\underline{x}'_r|$. Furthermore, (28) is an even function of the phase difference $\phi_{rr'}$, and therefore $P_{S,S',e}$ is independent of S and S'.

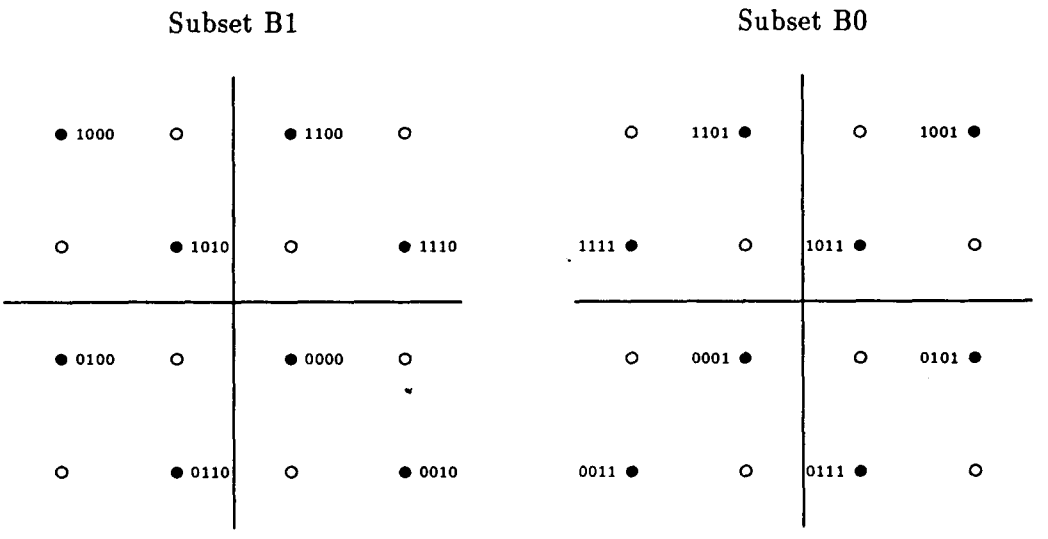


Figure 3: The first level of set partitioning of 16 QAM.

For $e \in \{xxx1\}$, where x is an arbitrary bit, the trellis states (S, S') must either $\in B1 \times B0$ or $\in B0 \times B1$. Again due to the special assignment of signal selectors, all distances generated between $B1$ and $B0$ are simply reflected in $B0$ and $B1$, which leaves the mutual angle magnitudes and vector lengths unaffected. Thus 16-QAM with the signal assignment of Figure 3 is quasi-regular with respect to the generalized distance of (28). Q.E.D.

Figure 4 shows the generalized distance spectrum of Ungerboeck's 8-PSK code $u4$ for several values ρ. It is worth noting that the spectrum not only shifts, but that the lines are also shifted with respect to each other. Because the generalized distance is reciprocal for constant envelope schemes, no new lines show up. Note also that the minimum generalized distance decreases as ρ increases.

Figure 5 shows the generalized distance spectrum of Ungerboeck's 16-QAM code $uq3$. Unlike the 8-PSK code, each original spectral line breaks up into a set of lines due to the fact that the signals are no longer reciprocal. For $\rho = 0.01$, these clusters of lines can still be identified with the original spectral line, while for larger values of ρ a more continuous spectrum emerges. It becomes evident that the minimum generalized distance $d(\mathbf{x}, \mathbf{x}')$, while indicative of a code's performance, is no longer the paramount performance parameter, but that a number of spectral lines influence the event error probability of the code.

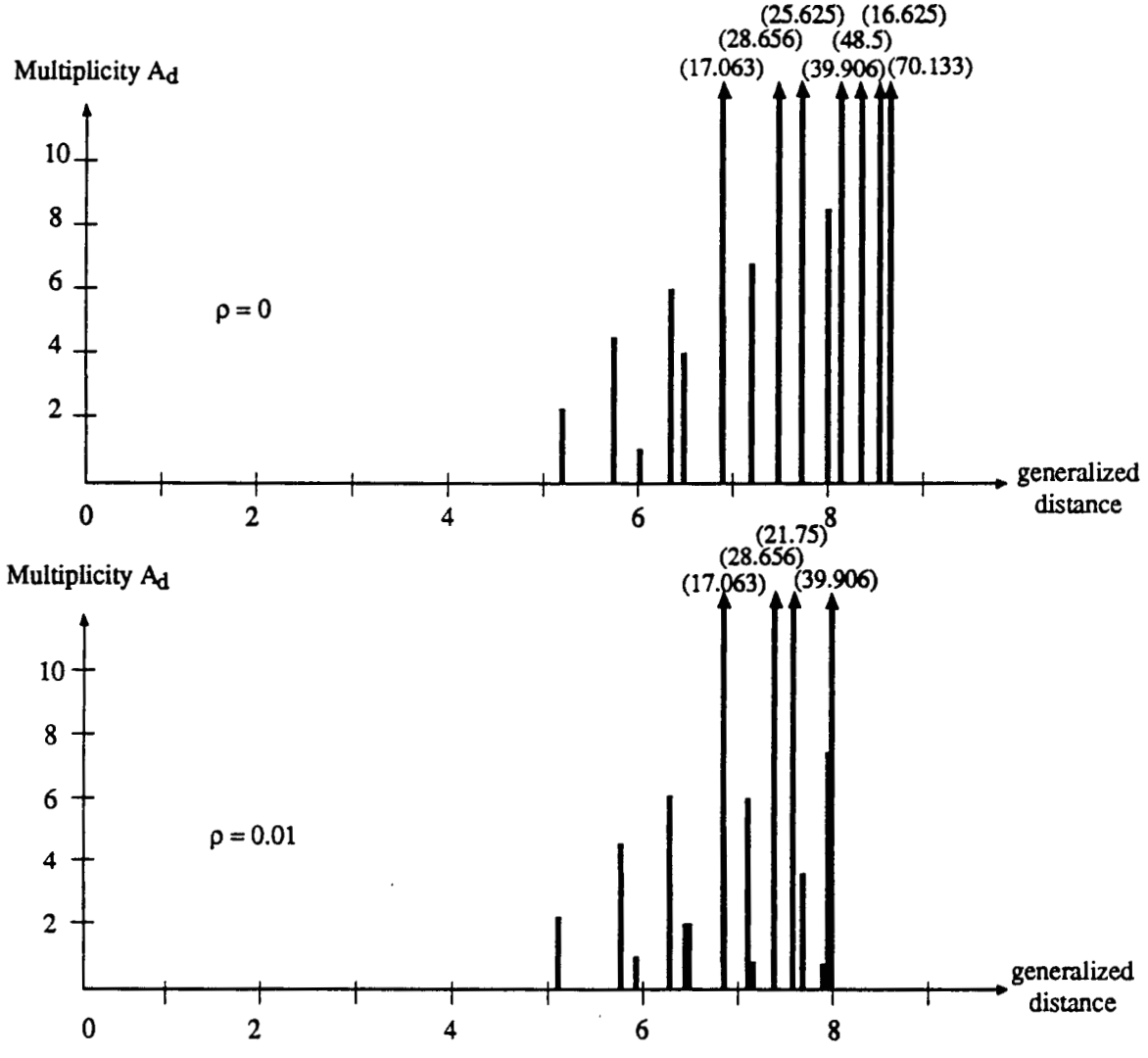


Figure 4.1: Spectrum of the 8-PSK code u4 for 5 different values of the phase-noise-to-channel-noise ratio ρ.

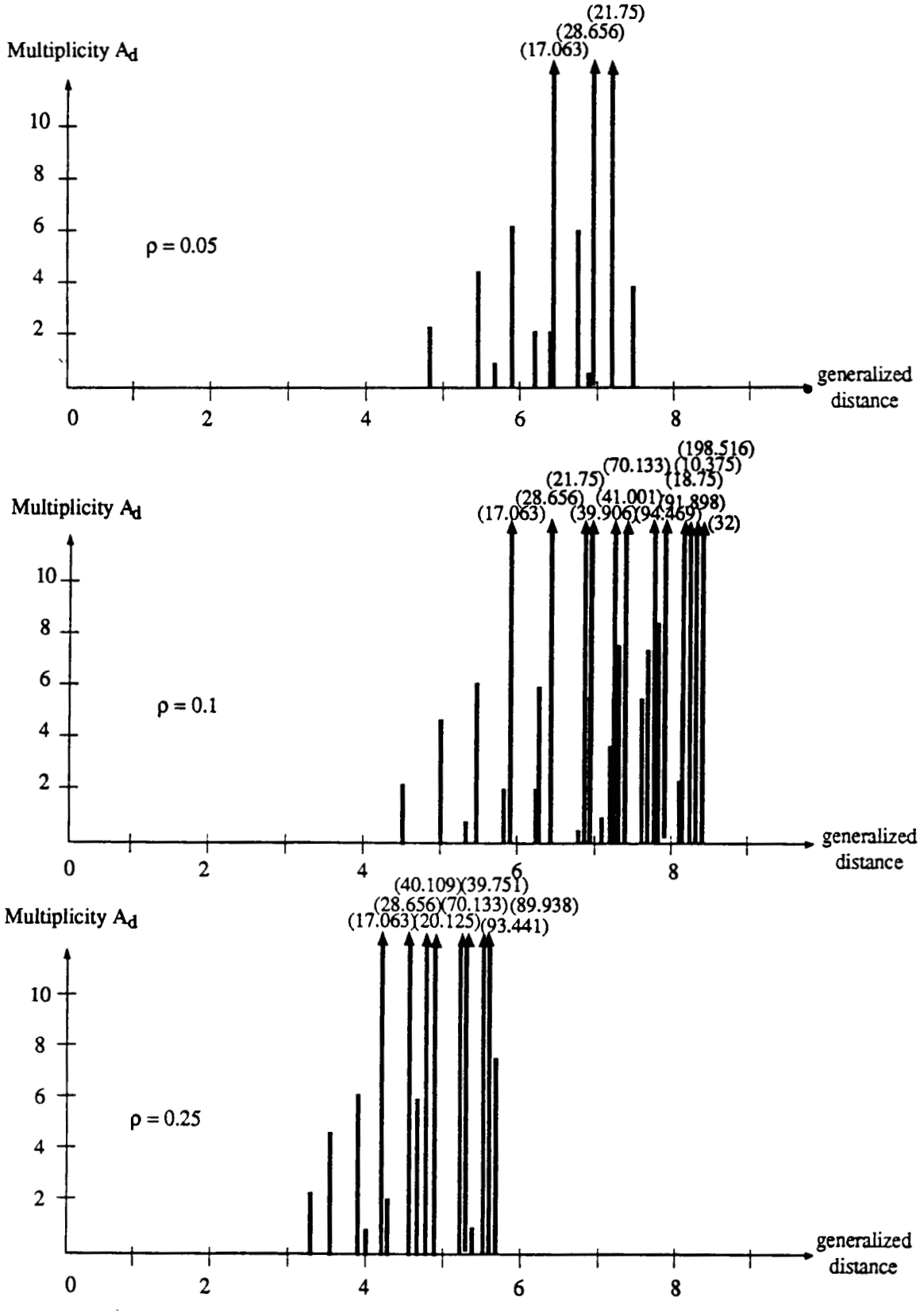


Figure 4.2: Spectrum of the 8-PSK code u4 for 5 different values of the phase-noise-to-channel-noise ratio ρ.

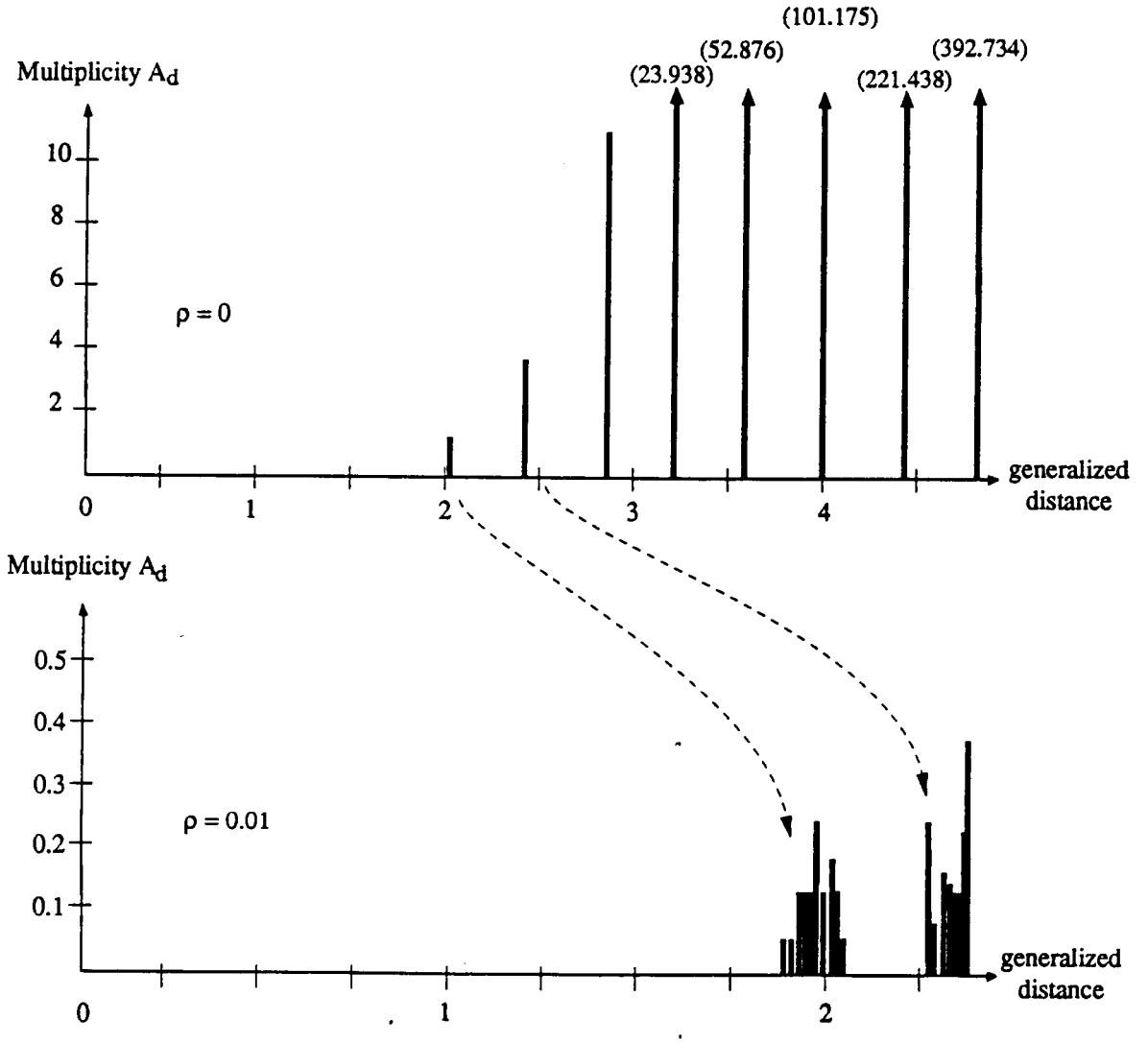


Figure 5.1: Spectrum of the 16-QAM code uq3 for 5 different values of the phase-noise-to-channel-noise ratio ρ.

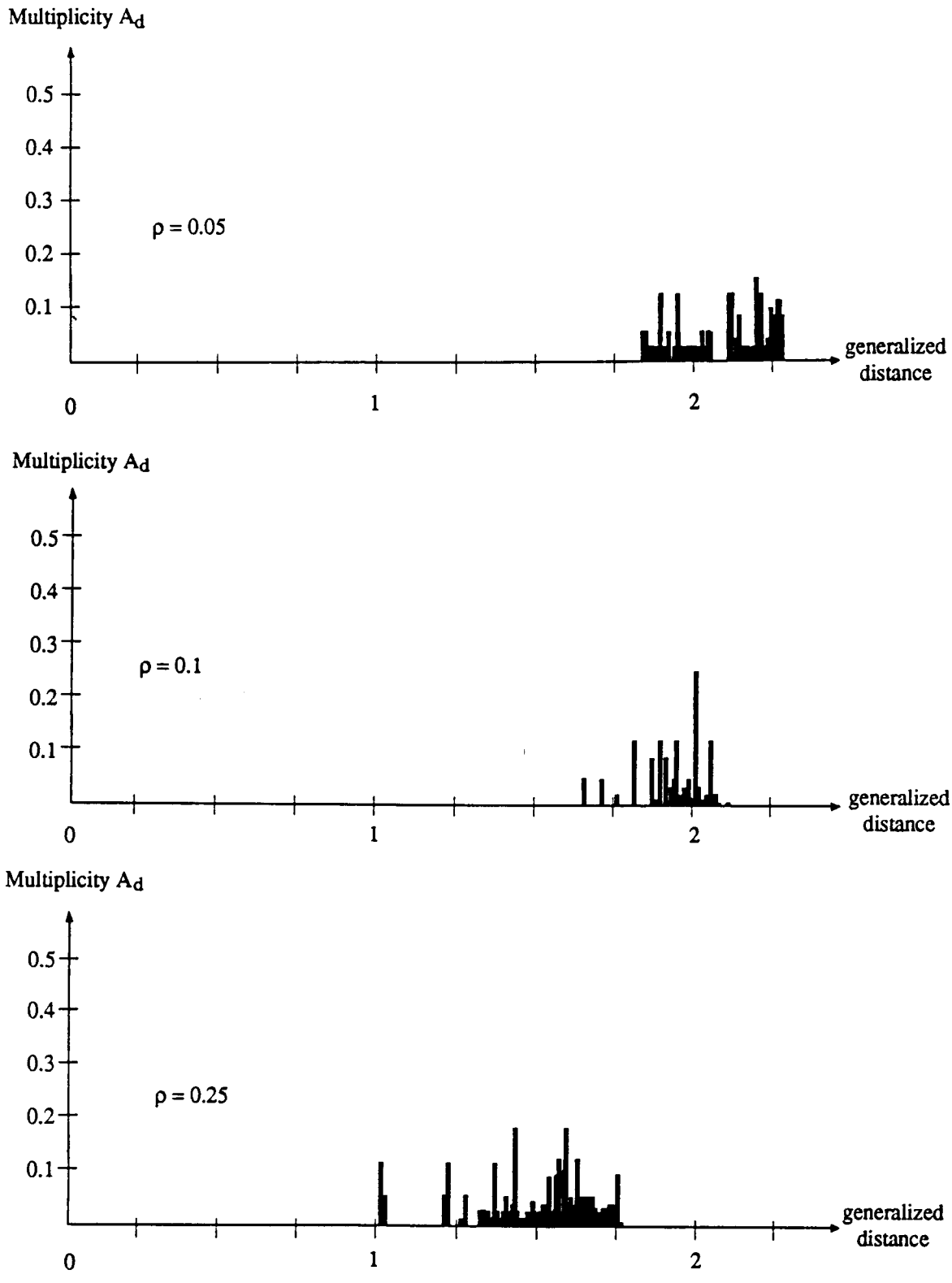


Figure 5.2: Spectrum of the 16-QAM code uq3 for 5 different values of the phase-noise-to-channel-noise ratio ρ.

Figure 6 shows the error performance of the code u4 with various phase-noise-to-channel-noise ratios. The event error probability of this code has been calculated using a variant of the distance spectrum algorithm for quasi-regular codes presented in [10]. It is evident that even though phase noise causes a degradation in system performance, this degradation does not affect the error probability dramatically. Even for phase-noise-to-channel-noise ratios as large as $\rho = 0.25$, the degradation is only 1.5 – 2dB. (Note that for a signal-to-noise ratio of 10dB, $\rho = 0.25$ translates into a phase standard deviation of $\sigma_\theta \approx 9°$.)

Figure 7 shows the error performance of the code u4 for several phase noise standard deviations σ_θ. Since it is unreasonable to assume that the phase jitter is independent of the channel noise, these curves do not represent actual system performance but are included to demonstrate that for phase errors $\sigma_\theta \lessapprox 5°$ the degradations due to phase jitter are small. Even at a target error rate of $P_e = 10^{-8}$ the degradation of a system with $5°$ phase jitter is less than $1dB$ compared with ideal synchronization. For larger values of the phase jitter (e.g., $\sigma_\theta = 10°$), the error curve flattens out at $P_e \approx 10^{-6}$, indicating a maximum achievable error performance for non-ideal phase synchronization.

Figure 8 shows the error performance of the code uq3 for various values of ρ. Most of the comments on 8-PSK coded modulation apply also to 16-QAM coded modulation. In general, the degradation suffered due to phase noise is larger however.

Figure 9 shows the increase in S/N required to achieve an event error probability of $P_e = 10^{-6}$ plotted versus σ_θ. The degradation curves for uncoded modulation are included for reference. It is worth noting that the expansion of the signal set which is introduced in TCM does not severely affect the degradation due to phase jitter, i.e., the degradation of uncoded QPSK is about the same as the degradation of coded 8-PSK for small phase jitter angles σ_θ. For larger phase jitter angles, coded 8-PSK is even more robust than uncoded QPSK and suffers significantly less degradation for $\sigma_\theta > 5°$. Note, however, that the degradation suffered by uncoded 8-PSK is much more severe.

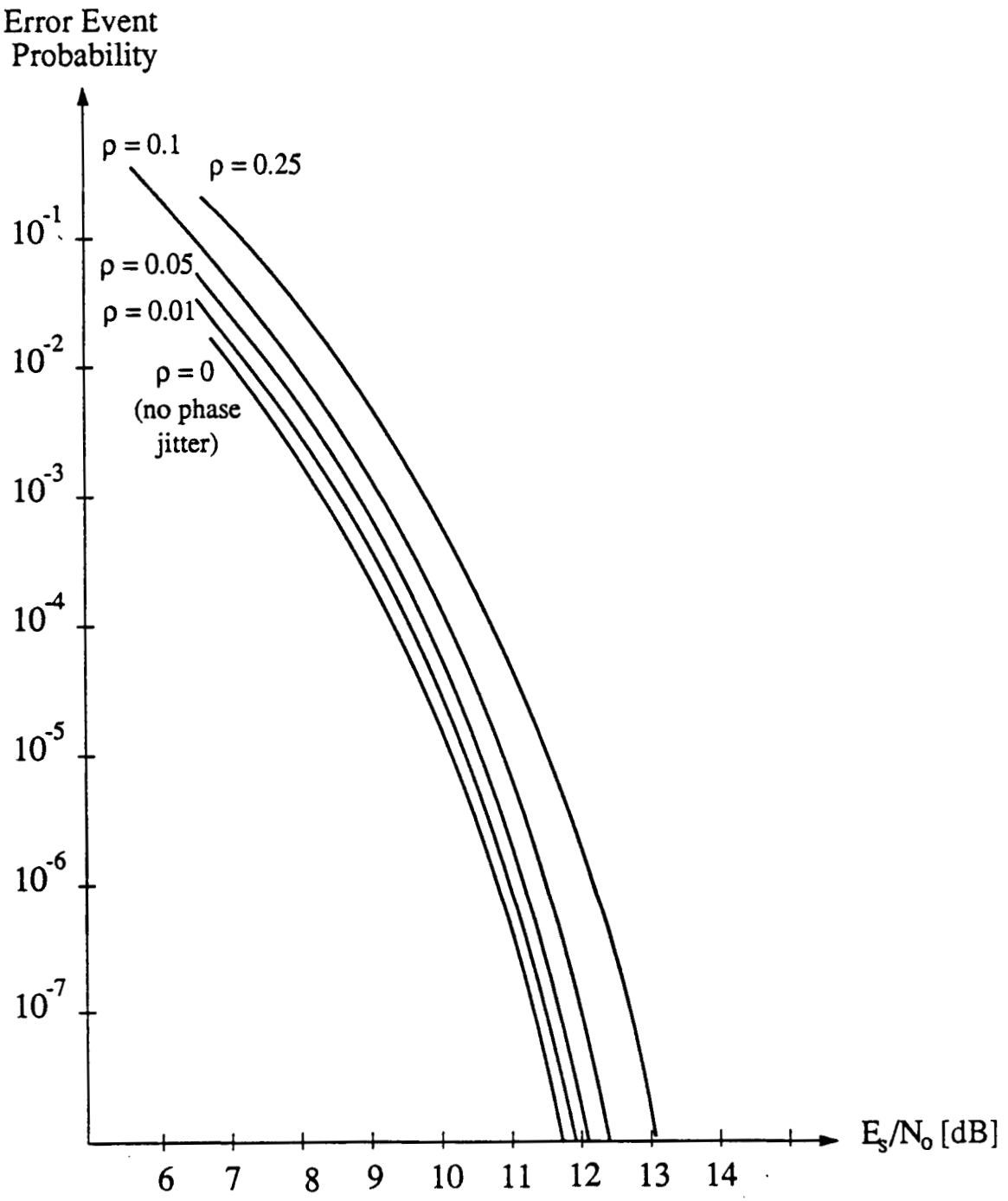


Figure 6: Event error probability of coded 8-PSK modulation (code u4) for various values of the phase-noise-to-channel-noise ratio.

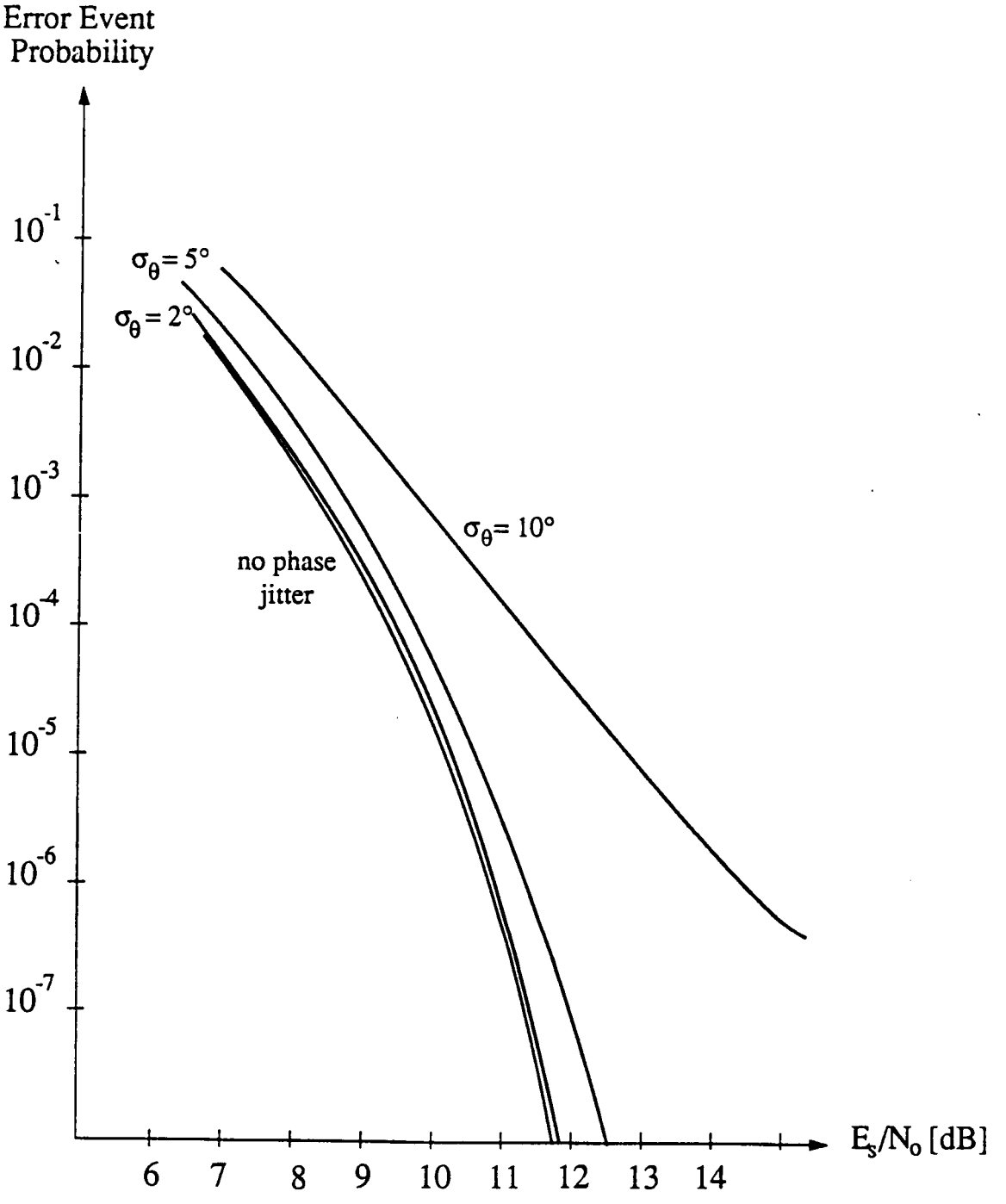


Figure 7: Event error probability of coded 8-PSK modulation (code u4) for various values of the phase angle standard deviation σ_θ.

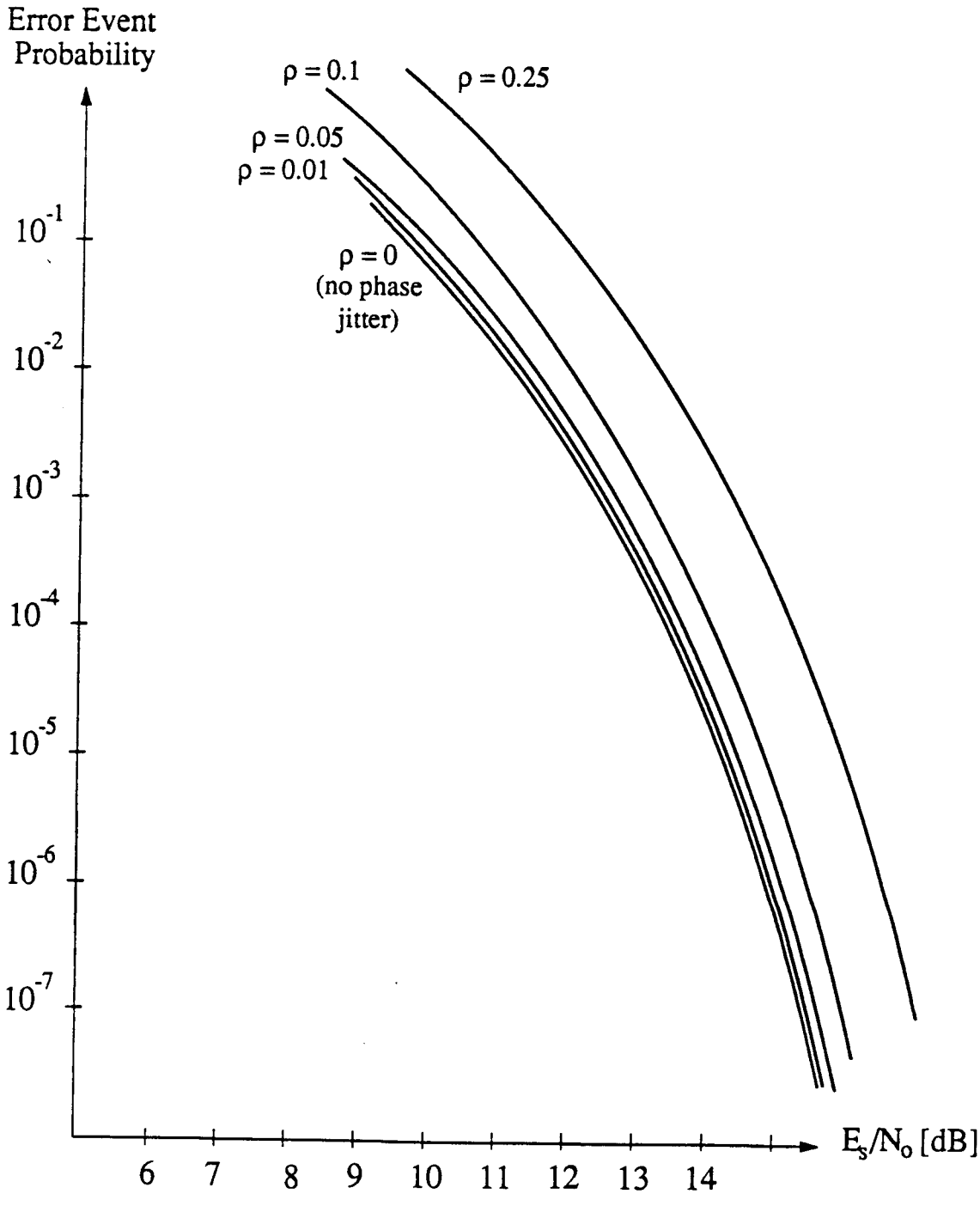


Figure 8: Event error probability of coded 16-QAM modulation (code uq3) for various values of the phase-noise-to-channel-noise ratio ρ.

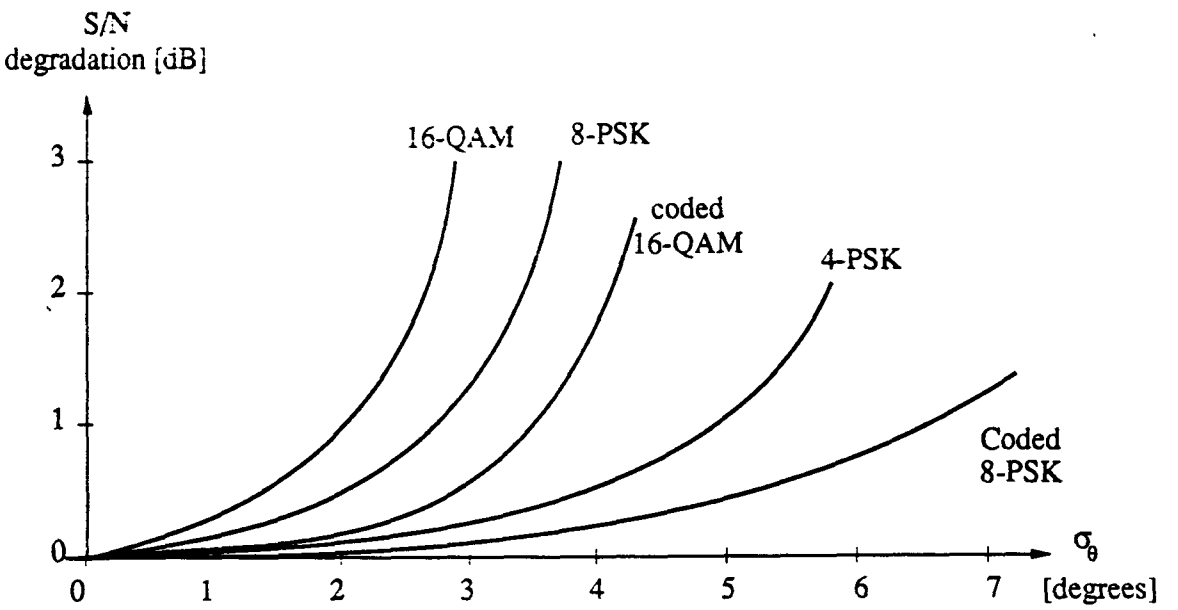


Figure 9: Signal-to-noise ratio penalties due to carrier phase jitter at an event error probability of $P_e = 10^{-6}$ for coded 8-PSK and 16-QAM modulation.

5 Carrier Phase Noise

In this section we show that the general method of analysis can be expanded to include other phenomena than channel noise which result in carrier phase jitter. As an example we discuss oscillator phase noise, which results because the receiver and transmitter oscillators will not be perfectly stable, but will exhibit phase fluctuations. These fluctuations result in *carrier phase noise* and are the principal cause of *line broadening.* They can be modeled as a white noise process [11]. The carrier is then $\cos(\omega_0 t + Z(t))$, where $Z(t)$ is an integrated white Gaussian noise process, i.e., the frequency noise $\dot{Z}(t)$ is closely modeled as white Gaussian noise with one-sided noise power spectral density N_z. The phase noise itself is

represented as

$$Z(t) = 2\pi \int_{\infty}^{t} \dot{Z}(\tau) d\tau. \tag{30}$$

Denote the Fourier transform of $Z(t)$ by $\mathcal{F}(Z(t)) = \mathcal{Z}(\omega)$. It can be calculated from the Fourier transform of $\mathcal{F}(\dot{Z}(t)) = \sqrt{N_z}$ as

$$\mathcal{Z}(\omega) = 2\pi \left(\frac{\sqrt{N_z}}{j\omega} + \pi \dot{Z}(0)\delta(\omega) \right), \tag{31}$$

where we may discard the DC-component $\pi \dot{Z}(0)\delta(\omega)$ since it is tracked by the phase-locked loop, and we obtain for the power spectrum of $Z(t)$

$$S_z(\omega) = |\mathcal{Z}(\omega)|^2 = 4\pi^2 \frac{N_z}{\omega^2}. \tag{32}$$

From (32) we see that the phase noise $Z(t)$ is a correlated Gaussian process. The phase noise variance σ_z^2 that results from the carrier frequency fluctuations can then be evaluated as [4, p. 27]

$$\sigma_z^2 = \frac{1}{2\pi} \int_{2\pi B_L}^{\infty} \frac{4\pi^2 N_z}{\omega^2} d\omega = \frac{N_z}{B_L}, \tag{33}$$

where B_L is the phase-locked loop bandwidth encountered earlier and we have assumed an ideal low-pass filter in the phase locked loop with bandwidth B_L.

Using a linear model for the phase locked loop, the two independent Gaussian contributions to the carrier phase jitter process result in a total Gaussian process, and we obtain for the carrier phase variance σ_θ^2

$$\begin{aligned} \sigma_\theta^2 &= \rho N_0 + \sigma_z^2 \\ &= B_L N_0 + \frac{N_z}{B_L} \\ &= \rho N_0 + \frac{N_z}{\rho}. \end{aligned} \tag{34}$$

It is evident that increasing the loop bandwidth B_L decreases the variance contribution σ_z^2 from the carrier phase noise because higher frequencies can be tracked, while it increases

the portion of the variance induced by channel noise. Decreasing the loop bandwidth has the opposite effect.

Equation (34) for the phase noise variance may now be used in (18) to evaluate the Chernoff factor, and we obtain

$$\begin{aligned} C(\underline{x}_r, \underline{x}'_r, \lambda) &= e^{-\lambda(1-\lambda N_0)|\underline{x}_r - \underline{x}'_r|^2} \frac{e^{\frac{2\lambda^2\sigma_\theta^2\kappa_r^2}{1-2\lambda\sigma_\theta^2\eta_r}}}{\sqrt{1-2\lambda\sigma_\theta^2\eta_r}} \\ &= e^{-\lambda(1-\lambda N_0)|\underline{x}_r - \underline{x}'_r|^2} \frac{e^{\frac{2\lambda^2(\rho N_0 + N_z/\rho)\kappa_r^2}{1-2\lambda(\rho N_0 + N_z/\rho)\eta_r}}}{\sqrt{1-2\lambda(\rho N_0 + N_z/\rho)\eta_r}}. \end{aligned} \tag{35}$$

This expression for the Chernoff factor may be used in the algorithm described in [10] with $\lambda = \lambda_{R_0}$ to evaluate the event error probability P_e of TCM schemes. Figure 10 shows the error performance of the code u4 with a noisy carrier and $N_z = 1.5 \times 10^{-4}[\text{rad}^2/s]$. Note that the error curves for different values of ρ cross. From (34) it can be seen that choosing a small ρ suppresses channel noise induced phase jitter while increasing the portion induced by phase noise. Thus for low signal-to-noise ratios a receiver with $\rho = 0.01$ outperforms a receiver with $\rho = 0.25$, while for high signal-to-noise ratios, where the phase noise portion is dominant, the situation is reversed, leading to the crossing of the error curves.

Figure 11 shows the error performance of code u4 for a high value of the carrier phase noise, i.e., $N_z = 2.5 \times 10^{-3}$. The curves for $\rho = 0.01$ and $\rho = 0.05$ demonstrate that there exists an irreducible error probability which is due to the phase noise of the carrier and which does not decrease with the channel S/N. This is manifested in the error curves' flattening out for large values of E_S/N_0, while the performance curves for no carrier phase noise show an exponential decrease with E_S/N_0 (Figure 6).

The results in this section are of an approximate and illustrative nature. More precise expressions for the phase jitter variance in (34) should be substituted for a detailed system performance analysis.

6 Conclusions

We have analyzed the performance of TCM schemes in the presence of phase jitter. The performance degradation due to phase jitter was calculated using the generalized distance spectrum of a code. In the case of practical coherence, when the phase jitter is induced only

by channel noise, the performance degradation was found to be moderate, even for relatively large values of the phase-noise-to-channel-noise ratio ρ. Other sources of phase jitter, such as oscillator phase noise, were found to result in more severe performance degradation.

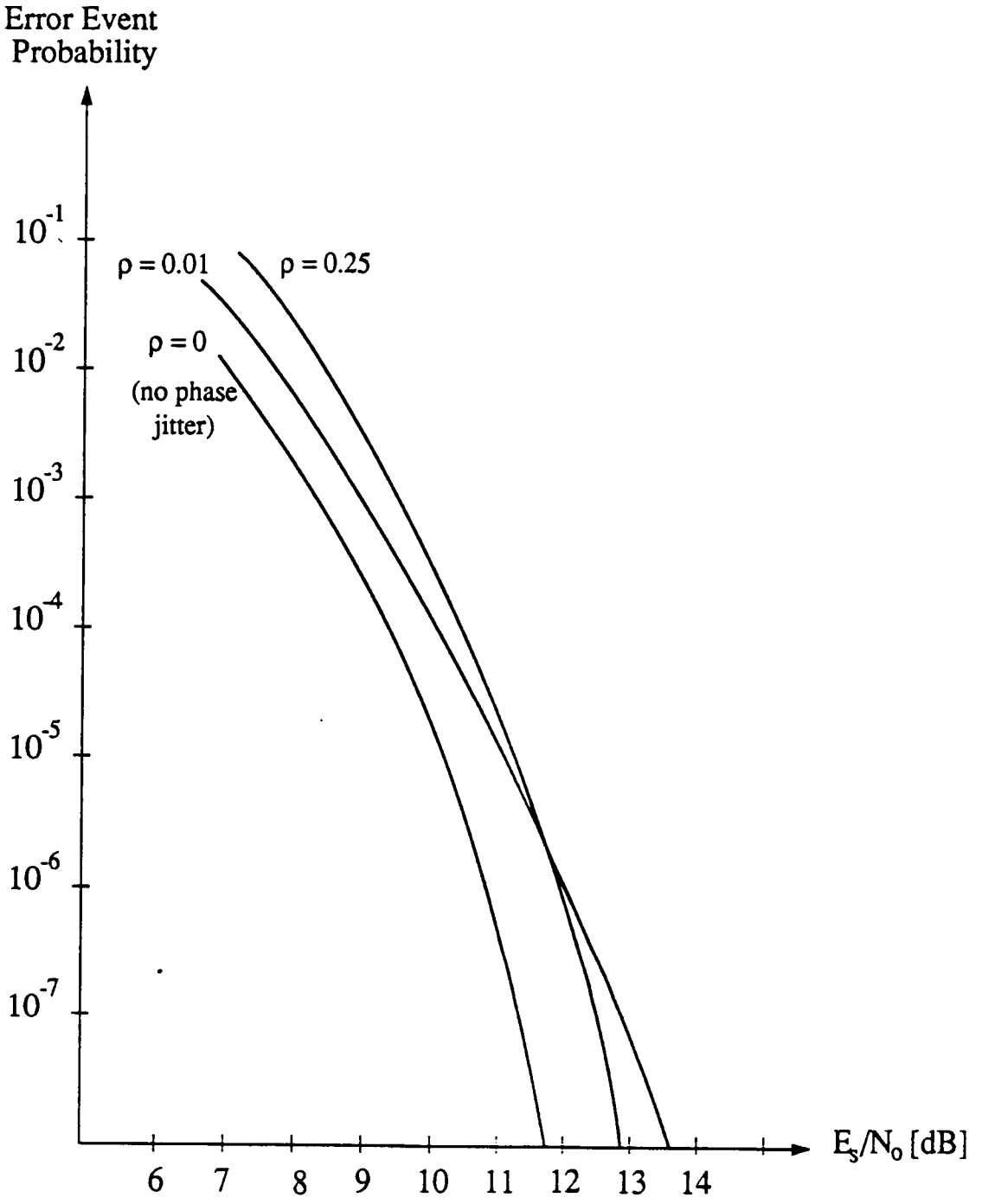


Figure 10: Event error probability of coded 8-PSK modulation (code u4) for various values of the phase-noise-to-channel-noise ratio ρ in the presence of carrier phase noise with value $N_z = 1.5 \times 10^{-4}$.

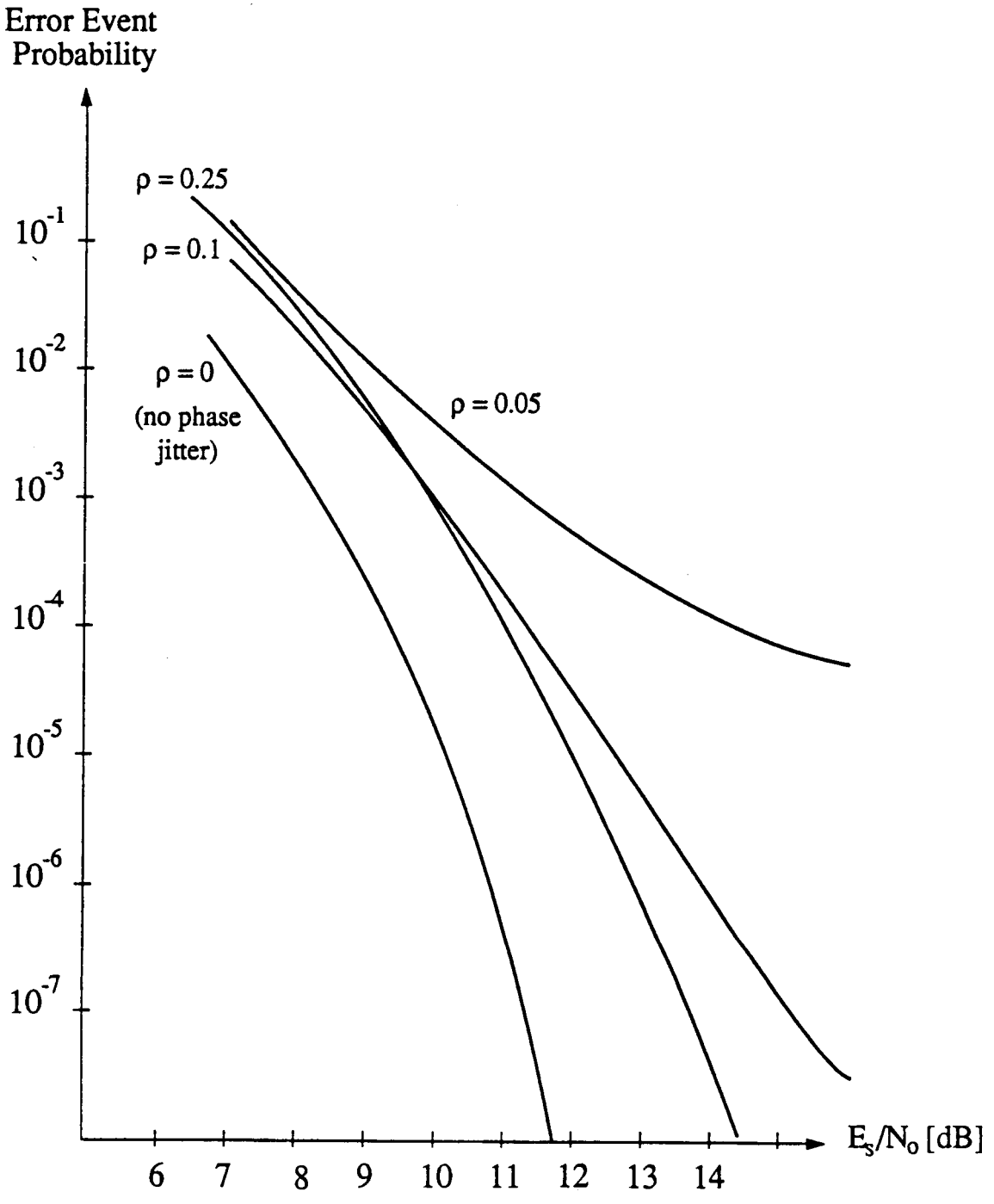


Figure 11: Event error probability of coded 8-PSK modulation (code u4) for various values of the phase-noise-to-channel-noise ratio ρ in the presence of carrier phase noise with value $N_z = 2.5 \times 10^{-3}$.

Appendix A

In this appendix we present the use of an adaptive phase predictor to whiten the residual phase errors. The carrier is tracked by a carrier recovery loop which produces the local reference signal $\cos(\omega_0 t + \theta(t))$, where $\theta(t)$ is the carrier recovery loop phase error. At the sampling instants nT, this phase error will give rise to phase offset angles θ_r. These phase offset angles are approximated by a Gaussian process [2,4]. Since the recovery loop has a small bandwidth, the phase error samples will be strongly correlated.

The method of achieving independent phase offset angles presented in this appendix is linear prediction. We will use a linear predictor driven by N previous symbol decisions to predict the phase offset θ_r of the r-th sampling instant, as illustrated in Figure A1. The N-th order predictor is fed by the measured phase offset θ'_j of the j-th received symbol $\underline{r}_j$ with respect to the j-th decoded symbol $\hat{\underline{a}}_j$, where $\hat{\underline{a}}_j$ is the best estimate of the j-th transmitted symbol $\underline{a}_j$ which is taken from the survivor with the best metric. The r-th phase offset θ'_r is then linearly predicted as

$$\hat{\theta}'_r = \sum_{j=1}^{N} h_j \theta'_{r-j}. \tag{A.1}$$

Note that the measured phase offset θ'_j is given by

$$\theta'_j = \theta_j + \beta_j, \tag{A.2}$$

where θ_j is the j-th phase offset of the recovery loop and β_j is a phase offset induced by the additive Gaussian noise affecting the j-th received signal point $\underline{y}_j$, as shown in Figure A2.

We are using the predicted phase offset $\hat{\theta}'_r$ at time unit r to compensate for the actual offset θ_r. The received signal point $\underline{y}_r$ is digitally rotated by $-\hat{\theta}'_r$ into $\underline{y}^*_r$.

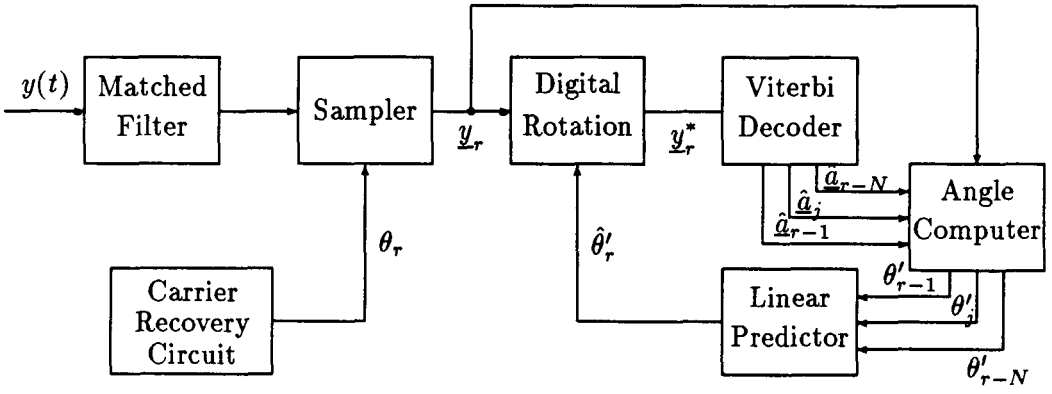


Figure A1: Phase recovery circuit with phase predictor.

The offset angles β_j are independent because they are caused by channel noise, which is white, and we may evaluate the autocorrelation function of θ'_j as

$$\begin{aligned} E\left[\theta'_j\theta'_{j-k}\right] &= E\left[(\theta_j+\beta_j)(\theta_{j-k}+\beta_{j-k})\right] \\ &= E\left[\theta_j\theta_{j-k}\right] + E\left[\beta_j\beta_{j-k}\right] \\ &= R_{\theta\theta}(k) + \delta(k)N_\beta, \end{aligned} \tag{A.3}$$

where $R_{\theta\theta}(k)$ is the autocorrelation function of the phase errors θ_j due to the phase recovery loop and $\delta(k)$ is the Dirac delta function. The cross terms in the above equation vanish due to the assumed independence of the offset angles β_j and the phase errors θ_j. A simple geometric consideration shows that $N_\beta \approx \frac{E_s}{N_0}$, where E_S is the signal energy (see [4]).

Figure A2: Combined effect of phase jitter and channel noise on the received signal point $\underline{y}_j$.

The linear predictor (A1) will predict θ'_τ using the optimum predictor coefficients $\mathbf{h} = [h_1, h_2, \cdots, h_N]^T$, which are given as [6]

$$\mathbf{h} = \mathbf{R}^{-1}\mathbf{r}, \tag{A.4}$$

where

$$\mathbf{R} = \begin{bmatrix} R_{\theta\theta}(0) + N_\beta & R_{\theta\theta}(1) & \cdots & R_{\theta\theta}(N-1) \\ R_{\theta\theta}(1) & R_{\theta\theta}(0) + N_\beta & \cdots & R_{\theta\theta}(N-2) \\ R_{\theta\theta}(2) & R_{\theta\theta}(1) & \cdots & R_{\theta\theta}(N-3) \\ \vdots & & & \vdots \\ R_{\theta\theta}(N-1) & R_{\theta\theta}(N-2) & \cdots & R_{\theta\theta}(0) + N_\beta \end{bmatrix} \tag{A.5}$$

is the correlation matrix of the phase error process θ'_j and $\mathbf{r} = [R_{\theta\theta}(1), R_{\theta\theta}(2), \cdots, R_{\theta\theta}(N)]^T$. A more detailed discussion on linear predictors is contained in [6].

The digital rotation of $\underline{y}_\tau$ leaves us with a residual phase error φ_τ. We are then interested in the autocorrelation function of φ_τ. Let us first consider the autocorrelation function of the total phase error process, i.e., of the difference between the full phase error θ'_j and the predicted phase error $\hat{\theta}'_j$, which is calculated as

$$
\begin{aligned}
&E\left[(\theta_r - \hat{\theta}'_r)(\theta_{r-k} - \hat{\theta}'_{r-k})\right] = \\
&= E\left[\left((\theta_r + \beta_r) - \sum_{j=1}^{N} h_j(\theta_{r-j} + \beta_{r-j})\right)\left((\theta_{r-k} + \beta_{r-k}) - \sum_{i=1}^{N} h_i(\theta_{r-i-k} + \beta_{r-i-k})\right)\right] \\
&= E\left[\theta_r\theta_{r-k}\right] + E\left[\beta_r\beta_{r-k}\right] - \sum_{i=1}^{N} h_i E\left[(\theta_r + \beta_r)\theta_{r-i-k}\right] - \sum_{i=1}^{N} h_i E\left[(\theta_r + \beta_r)\beta_{r-i-k}\right] \\
&\quad - \sum_{j=1}^{N} h_j E\left[\theta_{r-j}(\theta_{r-k} + \beta_{r-k})\right] - \sum_{j=1}^{N} h_j E\left[\beta_{r-j}(\theta_{r-k} + \beta_{r-k})\right] + \\
&\quad \sum_{j=1}^{N}\sum_{i=1}^{N} h_j h_i E\left[(\theta_{r-j} + \beta_{r-j})(\theta_{r-i-k} + \beta_{r-i-k})\right] \\
&= R_{\theta\theta}(k) + \delta(k)N_\beta - \sum_{i=1}^{N} h_i R_{\theta\theta}(i+k) - h_k N_\beta - \sum_{j=1}^{N} h_j R_{\theta\theta}(k-j) \\
&\quad + \sum_{i=1}^{N} h_i \sum_{j=1}^{N} h_j \left(R_{\theta\theta}(k+i-j) + \delta(k+i-j)\frac{N_\beta}{E_S}\right) \\
&= 0; \quad 1 \le |k| \le N. \qquad \text{(A.6)}
\end{aligned}
$$

The residual phase error process is a white Gaussian process with variance

$$
\begin{aligned}
E\left[(\theta'_r - \hat{\theta}'_r)(\theta'_r - \hat{\theta}'_r)\right] &= N_\beta + R_{\theta\theta}(0) - \sum_{j=1}^{N} h_j R_{\theta\theta}(j) \\
&= N_\beta + \sigma_\theta^2 - \sum_{j=1}^{N} h_j R_{\theta\theta}(j). \qquad \text{(A.7)}
\end{aligned}
$$

We are free to interpret the phase error process as the result of two independent processes. The first process is channel noise and results in a phase error with variance N_β. The second process is the residual phase error process with variance $\sigma_\theta^2 - \sum_{j=1}^{N} h_j R_{\theta\theta}(j)$. It is a white Gaussian process as required in section 1.

References

[1] T. Noguchi, Y. Daido and J. A. Nossek, "Modulation Techniques for Microwave Digital Radio", *IEEE Communications Magazine*, Vol. 24. No. 10, pp. 21–30, October 1986.

[2] W. C. Lindsey and M. K. Simon, *Telecommunication System Engineering*, Prentice-Hall, New Jersey, 1973.

[3] G. Ungerboeck, "Channel Coding with Multilevel/Phase Signals," *IEEE Trans. Inform. Theory*, Vol. IT-28, No.1, pp. 55–67, January, 1982.

[4] A. J. Viterbi, *Principles of Coherent Communications*, New York, McGraw-Hill, 1966.

[5] J. Hagenauer and C. W. Sundberg., "On the Performance Evaluation of Trellis-Coded 8-PSK Systems with Carrier Phase Offset," to be published in *Archiv für Elektronik und Übertragungstechnik.*, 1988.

[6] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, New Jersey, Prentice Hall, 1984.

[7] R. G. Gallager, *Information Theory and Reliable Communication*, New York, Wiley, 1968.

[8] C. Schlegel, "Bandwidth Efficient Coding for Non-Gaussian Channels", Ph. D. Dissertation, University of Notre Dame, Notre Dame, Indiana, 1988.

[9] H. Leib and S. Pasupathy, "Trellis-Coded MPSK with Reference Phase Errors", *IEEE Trans. Comm.*, Vol. COM-35, No.9, July 1987, pp. 888–900.

[10] M. Rouanne, D. J. Costello, Jr. and C. Schlegel, "An Algorithm for Computing the Distance Spectrum of Trellis Codes", Proceedings 1988 Conference on Information Sciences and Systems, Princeton, New Jersey, March 16–18, 1988.

[11] J. Salz, "Coherent Lightwave Communications", *AT&T Tech. J.*, Vol. 64, pp. 2153–2209, December 1983.

Appendix B

Bandwidth Efficient Coding on Channels Disturbed by Jamming and Impulse Noise

Bandwidth Efficient Coding on Channels Disturbed by Jamming and Impulse Noise [1]

Christian Schlegel

ABB Corporate Research
5405 Baden
Switzerland

Daniel J. Costello, Jr.

Department of Electrical and Computer Engineering
University of Notre Dame
Notre Dame, IN 46556

Abstract

Bandwidth efficient data transmission using trellis coded modulation (TCM) can significantly increase a digital communication link's reliability without sacrificing spectral efficiency. In this paper we discuss TCM on channels with impulse noise. The intentional partial band jammer with optimized duty cycle is analyzed analytically and taken as a worst case impulse noise interferer. It is shown that a new code design criterion, the effective length of a code becomes the central code parameter determining the performance of TCM. A number of new TCM codes are presented and their performance is calculated. It is shown that these codes fare significantly better than codes designed for additive white Gaussian noise channels.

1 Introduction

Practical communication channels are often not well modelled by *additive white Gaussian noise* (AWGN) channels. Yet most performance analyses of trellis coded modulation (TCM) have focussed on AWGN channels. In this paper we discuss a general class of impulse noise channels. Such impulse noise channels arise in various situation in practice. Man made environmental noise such as that caused by automotive ignition, neon lights or the switching of power machinery is impulsive. Intentional pulsed jamming, which is taken as the worst case impulse interference, also generates nonuniform noise. The most interesting application, however, is multiple access of a particular frequency band (FDMA) as envisioned in modern cellular radio. One proposed mechanism uses frequency hopping to overcome the often poor transmission quality of mobile radio channels. In such a situation, the probability that a particular user jumps to a frequency also claimed by another user is not zero. This is called a hit, and we denote the probability of a hit by ρ. An analogous situation arises in the case of time domain multiple access (TDMA).

As worst case situation we may view the interfering simultaneous users of a frequency band as a jammer trying to disrupt communication. We assume throughout the derivation that the channel symbols are interleaved, so as to assure independence of the jammer state at each time interval. The jammer shall be average power limited at the jamming power P_J. In a partial band jamming situation (FDMA), this jamming power is spread over a portion ρ of the total frequency band at a power spectral density N_J/ρ, where N_J is the power spectral density if the total jamming power P_J is spread over the entire frequency band. In a TDMA situation, the jammer will be pulsed during a fraction ρ of the time at a power spectral density N_J/ρ. In either case, for a user the jammer assumes two distinct modes of operation, it is on during a fraction ρ of the time at a power level N_J/ρ, and off the rest of the time $1-\rho$, where $\rho \in [0,1]$ is the jammer duty cycle. We further assume at this point that the decoder is furnished with perfect side information, i.e., at each symbol interval τ the jammer state, $s_\tau \in \{\text{on,off}\}$, is known.

A TCM communication system (Figure 1) consists of a trellis encoder, a signal interleaver, the transmission channel, a signal deinterleaver, and a trellis decoder. A rate $R = k/n$ trellis code is generated by a binary convolutional encoder followed by a mapper. The convolutional encoder is a finite state automaton

[1] This work was supported by NASA Grant NAG 5-557.

with 2^{ν} possible states, where ν is the memory order of the encoder. At each time interval r, the encoder accepts $\bar{k}$ binary input bits $(u_r^{\bar{k}}, u_r^{\bar{k}-1}, \ldots, u_r^1)$ and makes a transition from its state S_r at time r to one of $2^{\bar{k}}$ possible successor states S_{r+1}. The $\bar{n} = n-(k-\bar{k})$ output bits of the convolutional encoder and the $k-\bar{k}$ uncoded information bits $(u_r^k, \ldots, u_r^{\bar{k}+1})$ form one of 2^n binary n-tuples $v_r = (v_r^n, v_r^{n-1}, \ldots, v_r^1)$, which is translated by the mapper into one of $A = 2^n$ channel signals from a signal set $\mathcal{A} = \{\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_A\}$. The uncoded information bits do not affect the state of the convolutional encoder and cause $2^{k-\bar{k}}$ parallel transitions between the encoder states S_r and S_{r+1}. A rate $R = k/n$ trellis code transmits k bits/channel signal, where the channel signal set contains $A = 2^n$ signals. If such a TCM communication system replaces an uncoded system that uses a signal set with $A' = 2^k$ signals, the overall transmission rate is preserved, and we call such a TCM system bandwidth efficient. Due to their practical importance, we restrict attention to trellis codes for 2-dimensional signal sets.

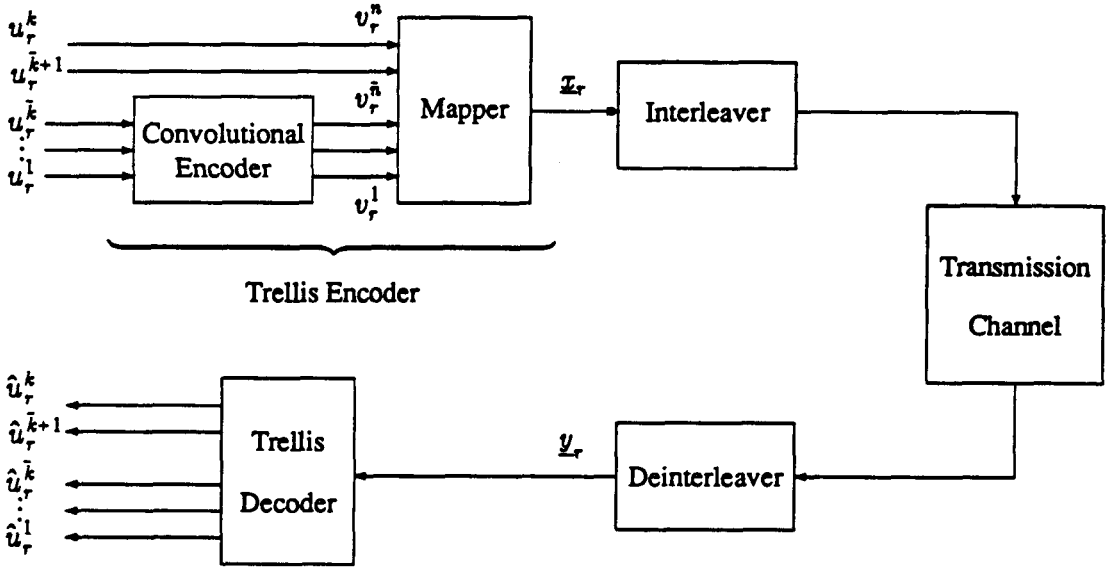


Figure 1: Trellis coded communication system.

The interleaver/deinterleaver converts the channel to a memoryless channel and insures that the signals in the received sequence are independent. They are decoded by a sequence estimator (usually using the Viterbi algorithm). The Viterbi algorithm finds the signal sequence that most closely corresponds to the sequence of received signals. It achieves this by calculating a decoding metric $m(\mathbf{x}, \mathbf{y})$ between $\mathbf{x}$ and $\mathbf{y}$, where $\mathbf{x} = (\underline{x}_1, \ldots, \underline{x}_l)$ is a possible sequence of 2-dimensional transmitted signals and $\mathbf{y} = (\underline{y}_1, \ldots, \underline{y}_l)$ is the received sequence. $m(\mathbf{x}, \mathbf{y})$ is some non-negative function of $\mathbf{x}$ given $\mathbf{y}$, which is inversely related to the conditional probability that $\mathbf{x}$ was transmitted if $\mathbf{y}$ was received. The decoder will then choose the message sequence $\mathbf{x}$ for which this metric is minimized. It makes an error if it decodes a sequence $\mathbf{x}'$, given that the correct sequence, i.e., the transmitted sequence, was $\mathbf{x}$. This will happen if $m(\mathbf{x}', \mathbf{y}) \leq m(\mathbf{x}, \mathbf{y})$. In the case of channel state side information the decoder will use $m(\mathbf{x}, \mathbf{b}, \mathbf{y})$ as its metric, where $\mathbf{b}$ is the side information obtained from the channel.

2 Optimum Receiver for Impulse Noise Channels

In this section we will discuss the optimum receiver strategy if the channel's noise disturbance can be modelled as impulse noise as defined in the introduction. In [Wozencraft-Jacobs, 1965], the optimum receiver is defined as the receiver using the *maximum a posteriori* (MAP) probabilities as the decision criteria. In the

case of equally likely messages – the case we will discuss – this optimum receiver becomes the *maximum-likelihood* (ML) receiver. The MAP-receiver chooses the message $\mathbf{x}_k$ if and only if

$$\Pr(\mathbf{x}_k \mid \mathbf{y}) \leq \Pr(\mathbf{x}_i \mid \mathbf{y}); \quad \text{for } i = 1, \cdots, M, i \neq k, \tag{1}$$

where M is the number of messages. Using Bayes' Rule, (1) is equivalent to choosing the message $\mathbf{x}_k$ for which $\Pr(\mathbf{y} \mid \mathbf{x}_k)$ is largest. As the transmitted signals are interleaved the noise in our channel is a non-uniform non-correlated Gaussian process and $\Pr(\mathbf{y} \mid \mathbf{x})$ is given by

$$\begin{aligned} \Pr(\mathbf{y} \mid \mathbf{x}) &= \prod_{\tau=0}^{l} \frac{1}{\sqrt{\pi N_{c\tau}}} e^{\frac{-|\underline{y}_\tau - \underline{x}_\tau|^2}{N_{c\tau}}} \\ &= \exp\left(-\sum_{\tau=0}^{l} \frac{|\underline{y}_\tau - \underline{x}_\tau|^2}{N_{c\tau}}\right) \prod_{\tau=0}^{l} \frac{1}{\sqrt{\pi N_{c\tau}}}, \end{aligned} \tag{2}$$

where the channel noise power spectral density $N_{c\tau}$ is given by

$$N_{c\tau} = \begin{cases} N_b = N_J/\rho + N_0 & \text{with probability } \rho \ (s_\tau = \text{on}) \\ N_g = N_0 & \text{with probability } 1 - \rho \ (s_\tau = \text{off}). \end{cases} \tag{3}$$

and N_0 is the channels thermal noise power spectral density. If $N_J = 0$ or $\rho = 1$, we have the ordinary AWGN channel. Selecting the message $\mathbf{x}_k$ which maximizes (2) is equivalent to selecting the message that minimizes the metric

$$m(\mathbf{x}, \mathbf{y}) = \sum_{\tau=0}^{l} \frac{N_0}{N_{c\tau}} |\underline{y}_\tau - \underline{x}_\tau|^2, \tag{4}$$

because $\prod_{\tau=0}^{l} \frac{1}{\sqrt{\pi N_{c\tau}}}$ and N_0 are independent of the particular message sent. The metric in (4) is the sum of individual branch metrics and the optimum receiver will employ the maximum-likelihood branch metric $c(s_\tau)|\underline{y}_\tau - \underline{x}_\tau|^2$, where

$$c(s_\tau) = \frac{N_0}{N_{c\tau}} = \begin{cases} \frac{N_0}{N_J/\rho + N_0} & \text{if } s_\tau = \text{on} \\ 1 & \text{if } s_\tau = \text{off}. \end{cases} \tag{5}$$

Note that if the jamming power spectral density $N_J >> N_0$, the receiver will give a very low metric to any symbol received while the jammer is active, that is it will mark those symbols as unreliable.

3 Error Bounds

Because it is unfeasible to exactly calculate the error performance even on a AWGN channel, we resort to bounding techniques. Denote the two code word error probability when $\mathbf{x}$ is sent and $\mathbf{x}'$ is decoded, given the jammer state sequence $\mathbf{s} = (s_1, s_2, \ldots, s_l)$ by $P(\mathbf{x} \to \mathbf{x}' \mid \mathbf{s})$.

The two code word error probability can then be expressed in terms of the Chernoff bound, i.e.,

$$\begin{aligned} P(\mathbf{x} \to \mathbf{x}' \mid \mathbf{s}) &= \underset{\mathbf{S}}{E}\left[P(\mathbf{x} \to \mathbf{x}')\right] \\ &\leq \underset{\mathbf{s}}{E}\left[\min_{\lambda} \underset{\mathbf{y}}{E}\left[\prod_{\tau=1}^{l} \exp\left(-\lambda c(s_\tau)\left(|\underline{y}_\tau - \underline{x}'_\tau|^2 - |\underline{y}_\tau - \underline{x}_\tau|^2\right)\right)\right]\right] \end{aligned}$$

$$\leq \min_{\lambda} \underset{\underline{s}}{E}\underset{\underline{y}}{E}\left[\prod_{r=1}^{l} \exp(-\lambda c(s_r)\left((|\underline{y}_r - \underline{x}'_r|^2 - |\underline{y}_r - \underline{x}_r|^2)\right)\right]$$

$$= \min_{\lambda} \underset{\underline{s}}{E}\underset{\underline{y}}{E}\left[\prod_{r=1}^{l} \exp\left(-\lambda c(s_r)\left((\underline{x}'^2_r - \underline{x}^2_r) + 2\,\underline{y}_r(\underline{x}_r - \underline{x}'_r)\right)\right)\right]$$

$$= \min_{\lambda} C(\mathbf{x}, \mathbf{x}', \lambda). \tag{6}$$

As the transmitted symbols are interleaved, the received symbols $\underline{y}_r$ are independent from one another. Noting that $\underline{y}_r = \underline{x}_r + \underline{n}_r$, where $\underline{n}_r$ is 2-dimensional Gaussian noise with one-sided noise power spectral density given by (3), and applying the second law of expectation to the above equation lets us move the average over y into the product, i.e.,

$$C(\mathbf{x}, \mathbf{x}', \lambda) = \underset{\underline{s}}{E}\left[\prod_{r=1}^{l} \exp(-\lambda c(s_r)(\underline{x}_r - \underline{x}'_r)^2) \underset{\underline{n}_{cr}}{E}\left\{\exp(-2\lambda c(s_r)\underline{n}_r(\underline{x}_r - \underline{x}'_r))\right\}\right]$$

$$= \underset{\underline{s}}{E}\left[\prod_{r=1}^{l} \exp(-\lambda c(s_r)(\underline{x}_r - \underline{x}'_r)^2) \exp(\lambda^2 c^2(s_r) N_{cr}(\underline{x}_r - \underline{x}'_r)^2)\right], \tag{7}$$

Rewriting the above equation as

$$C(\mathbf{x}, \mathbf{x}', \lambda) = \underset{\underline{s}}{E}\left[\prod_{r=1}^{l} \exp(-\lambda c(s_r)(\underline{x}_r - \underline{x}'_r)^2 (1 - \lambda c(s_r) N_{cr}))\right], \tag{8}$$

and observing that $c(s_r) N_{cr} = N_0$ is independent of the jammer state, we note that (8) is minimized independently of r by setting $\lambda = \lambda_{min} = 1/2\,N_0$, i.e.,

$$C(\mathbf{x}, \mathbf{x}', \lambda_{min}) \leq \underset{\underline{s}}{E}\left[\prod_{r=1}^{l} \exp\left(-\frac{(\underline{x}_r - \underline{x}'_r)^2}{4\,N_{cr}}\right)\right]. \tag{9}$$

This Chernoff bound breaks up into a product over the individual jammer states s_r, which we call *Chernoff factors*, i.e.,

$$C(\mathbf{x}, \mathbf{x}', \lambda_{min}) = \prod_{r=1}^{l} C(x_r, x'_r) = \prod_{r=1}^{l} \underset{s_r}{E}\left[\exp\left(-\frac{(\underline{x}_r - \underline{x}'_r)^2}{4\,N_{cr}}\right)\right]. \tag{10}$$

Due to the interleaving process, the jammer states s_r are statistically independent and we may average over s, that is, we may drop the dummy subscript r as the jammer states are independent identically distributed (i.i.d). (Note that without interleaving we could still proceed further analytically if the jammer state sequence can be described by a Markov chain, see [Gallager, 1968]). The Chernoff factors are then given by

$$C(\mathbf{x}, \mathbf{x}', \lambda_{min}) = \prod_{r=1}^{l} C(x_r, x'_r) = \prod_{r=1}^{l} \underset{\underline{s}}{E}\left[\exp\left(-\frac{(\underline{x}_r - \underline{x}'_r)^2}{N_{cr}}\right)\right]. \tag{11}$$

The cutoff-rate of the channel, R_0 can be calculated [Schlegel, 1988] as,

$$R_0 = -\log_2 \sum_{m=1}^{A}\sum_{p=1}^{A} p(\underline{a}_m)p(\underline{a}_p)C(\underline{a}_m, \underline{a}_p)$$

$$= -\log_2\left(\sum_{m=1}^{A}\sum_{p=1}^{A} p(\underline{a}_m)p(\underline{a}_p)\underset{\underline{s}}{E}\left[\exp(-\frac{(\underline{a}_m - \underline{a}_p)^2}{4\,N_{cr}})\right]\right), \tag{12}$$

where the Chernoff factor between $\underline{a}_m$ and $\underline{a}_p$, i.e.,

$$C(\underline{a}_m, \underline{a}_p) = \underset{s}{E}\left[\exp\left(-\frac{(\underline{a}_m - \underline{a}_p)^2}{4N_{cr}}\right)\right]$$

$$= \rho e^{-\frac{(\underline{a}_m - \underline{a}_p)^2}{4(N_0 + N_J/\rho)}} + (1-\rho)e^{-\frac{(\underline{a}_m - \underline{a}_p)^2}{4N_0}}. \quad (13)$$

An enemy jammer will try to minimize the cutoff rate by varying ρ, as this will raise the error probability on the average for all coding systems. Rather than minimizing (12) for one particular constellation we will introduce an upper bound on (12), which is exact for the simplex constellation, and then minimize this upper bound. This will provide a robust minimizing ρ which is valid for any constellation.

Let us now proceed by substituting (13) into (12) to obtain

$$R_0 = -\log_2\left\{\min_{p_i}\sum_{m=1}^{A}\sum_{p=1}^{A} p(\underline{a}_m)p(\underline{a}_p)\left[\rho e^{-\frac{(\underline{a}_m - \underline{a}_p)^2}{4(N_0 + N_J/\rho)}} + (1-\rho)e^{-\frac{(\underline{a}_m - \underline{a}_p)^2}{4N_0}}\right]\right\}. \quad (14)$$

Defining $b = \sum_{i=1}^{A} p(\underline{a}_i)^2$, we may rewrite the above expression as

$$2^{-R_0} = \min_{p_i}\left(b + (1-b)\sum_{\substack{m=1 \\ m\neq p}}^{A}\sum_{p=1}^{A}\frac{p(\underline{a}_m)p(\underline{a}_p)}{1-b}\left[\rho e^{-\frac{(\underline{a}_m - \underline{a}_p)^2}{4(N_0 + N_J/\rho)}} + (1-\rho)e^{-\frac{(\underline{a}_m - \underline{a}_p)^2}{4N_0}}\right]\right). \quad (15)$$

Introducing the factor $(1-b)$ in the second term above makes it an expectation. As the exponential function is $\cup$-convex we may apply Jensen's inequality to the second term to obtain

$$2^{-R_0} \geq \min_{p_i}\left(b + (1-b)\left[\rho\exp\left(-\frac{\sum_{m=1}^{A}\sum_{p=1}^{A} p(\underline{a}_m)p(\underline{a}_p)(\underline{a}_m - \underline{a}_p)^2}{4(N_0 + N_J/\rho)(1-b)}\right)\right.\right.$$

$$\left.\left. + (1-\rho)\exp\left(-\frac{\sum_{m=1}^{A}\sum_{p=1}^{A} p(\underline{a}_m)p(\underline{a}_p)(\underline{a}_m - \underline{a}_p)^2}{(4N_0)(1-b)}\right)\right]\right). \quad (16)$$

As an optimal signal set has its centroid at the origin, i.e., $\sum_{i=1}^{A} p(\underline{a}_i)\underline{a}_i = 0$, (see [Wozencraft-Jacobs, page 248] or [Massey, 1974]), we identify the above summations as

$$\sum_{m=1}^{A}\sum_{p=1}^{A} p(\underline{a}_m)p(\underline{a}_p)(\underline{a}_m - \underline{a}_p)^2 = 2E_S, \quad (17)$$

where $E_S = \underset{i}{E}[\underline{a}_i^2]$, the average signal energy. Substituting into (16) we obtain

$$2^{-R_0} \geq \min_{p_i}\left(b + (1-b)\left[\rho e^{-\frac{2E_S}{4(N_0 + N_J/\rho)(1-b)}} + (1-\rho)e^{-\frac{2E_S}{(4N_0)(1-b)}}\right]\right). \quad (18)$$

Using Lagrange's method it can be shown that the minimizing distribution in the above lower bound is the uniform distribution $p_i = 1/A$. Using this we may further simplify the bound to obtain

$$2^{-R_0} \geq \frac{1}{A} + \left(1 - \frac{1}{A}\right)\left(\rho e^{-\frac{2E_S}{4(N_0 + N_J/\rho)(1-1/A)}} + (1-\rho)e^{-\frac{2E_S}{(4N_0)(1-1/A)}}\right). \quad (19)$$

This equation is an upper bound for any constellation and equality holds for the simplex signal set for which $(\underline{a}_m - \underline{a}_p)^2$ is constant for all m, p such that $m \neq p$.

An enemy jammer will now try to minimize R_0 for any constellation. It will therefore maximize the right hand side of (19) and evaluate ρ such that

$$\frac{\partial}{\partial \rho}\left(\frac{1}{A}+\left(1-\frac{1}{A}\right)\left(\rho e^{-\frac{2E_S}{4(N_0+N_J/\rho)(1-1/A)}}+(1-\rho)e^{-\frac{2E_S}{(4N_0)(1-1/A)}}\right)\right)=$$

$$\left(1-\frac{1}{A}\right)\left\{e^{-\frac{2E_S}{4(N_0+N_J/\rho)(1-1/A)}}\left(1-\rho\frac{2E_SN_J}{4(\rho N_0+N_J)^2(1-1/A)}\right)\right.$$

$$\left.-e^{-\frac{2E_S}{4N_0(1-1/A)}}\right\}=0. \tag{20}$$

As the jammer will not have precise knowledge of the noise power density N_0 at the receiver, and as $N_J >> N_0$, it seems sensible to neglect N_0 in the above equation and set

$$e^{-\frac{2E_S}{4(N_0+N_J/\rho)(1-1/A)}}\left(1-\rho\frac{2E_SN_J}{4(\rho N_0+N_J)^2(1-1/A)}\right)-e^{-\frac{2E_S}{4N_0(1-1/A)}}$$

$$\approx e^{-\frac{\rho E_S}{2N_J(1-1/A)}}\left(1-\rho\frac{E_S}{2N_J(1-1/A)}\right)=0. \tag{21}$$

This maximum is achieved by setting

$$\rho=\begin{cases}\frac{2N_J(1-1/A)}{E_S} & \text{for } E_S/N_J > 2(1-1/A)\\ 1 & \text{otherwise.}\end{cases} \tag{22}$$

Equation (22) reveals that a smart jammer will shorten the jamming duty cycle as the signal energy increases in the transmission to be disturbed. Substituting the maximizing ρ into (13) we obtain for the Chernoff factor between $\underline{x}_r$ and $\underline{x}'_r$

$$C(\underline{x}_r,\underline{x}'_r)=\frac{2N_J(1-1/A)}{E_S}e^{-\frac{(\underline{x}_r-\underline{x}'_r)^2}{4N_0+\frac{4E_S}{2(1-1/A)}}}+\left(1-\frac{2N_J(1-1/A)}{E_S}\right)e^{-\frac{(\underline{x}_r-\underline{x}'_r)^2}{4N_0}}. \tag{23}$$

Even though it seems that this Chernoff factor exhibits an exponential characteristic, this is not the case, which is shown by the following argument. Let $d^2_{rr'}$ be the normalized distance between the signals $\underline{x}_r$ and $\underline{x}'_r$, i.e., $d^2_{rr'}$ is independent of the signal energy E_S and is defined as $d^2_{rr'} = (\underline{x}_r - \underline{x}'_r)^2/E_S$. We may now write

$$C(\underline{x}_r,\underline{x}'_r)=\frac{2N_J(1-1/A)}{E_S}e^{-\frac{d^2_{rr'}E_S}{4N_0+\frac{4E_S}{2(1-1/A)}}}+\left(1-\frac{2N_J(1-1/A)}{E_S}\right)e^{-\frac{d^2_{rr'}E_S}{4N_0}}. \tag{24}$$

As we are primarily dealing with a jammed channel (i.e., $N_J >> N_0$ as stipulated earlier), we will further assume that $E_S >> N_0$, as otherwise $E_S < N_J$ and then the transmitter would not be successful in sending any information. We now obtain

$$C(\underline{x}_r,\underline{x}'_r)\approx\frac{2N_J(1-1/A)}{E_Se^{d^2_{rr'}(1-1/A)/2}}+\left(1-\frac{2N_J(1-1/A)}{E_S}\right)e^{-\frac{d^2_{rr'}E_S}{4N_0}}. \tag{25}$$

Note that

$$C(\underline{x}_\tau, \underline{x}'_\tau) \approx \begin{cases} \frac{2(1-1/A)}{e^{\frac{d^2_{\tau\tau'}}{4N_0}(1-1/A)/2}} \frac{N_J}{E_S} = K_{\tau\tau'} \frac{N_J}{E_S} & \text{if } \underline{x}_\tau \neq \underline{x}'_\tau \\ 1 & \text{if } \underline{x}_\tau = \underline{x}'_\tau. \end{cases} \tag{26}$$

where $K_{\tau\tau'}$ is a constant that depends on the signals used, but not on the signal to noise ratio E_S/N_J.

It is interesting to note that the jammer achieves a degradation of the Chernoff bound in the sense that $C(\underline{x}_\tau, \underline{x}'_\tau)$ is no longer an exponential bound of the form $e^{-K_{\tau\tau'}\frac{E_S}{N_J}}$ but is only inversely proportional to the signal to noise ratio E_S/N_J as evidenced in (26).

Using (26) in (12) we obtain for the cutoff-rate R_0

$$\begin{aligned} R_0 &= -\log_2 \left(\sum_{k=1}^{A} \sum_{l=1}^{A} p_k p_l \frac{2N_J(1-1/A)}{E_S e^{d_{kl}(1-1/A)/2}} + \left(1 - \frac{2N_J(1-1/A)}{E_S}\right) e^{-\frac{d_{kl}E_S}{4N_0}} \right) \\ &\leq -\log_2 \left(\frac{1}{A} + \frac{2N_J}{E_S e} \right). \end{aligned} \tag{27}$$

The upper bound shows clearly that the error probability no longer decreases exponentially but only polynomially with the signal to noise ratio E_S/N_J. This will be of importance to the code design for jammed channels.

Figure 2 shows the cutoff rate for some popular signal constellations using the minimizing value of the duty cycle ρ, as well as the upper bound on R_0, i.e., (27) for $A = 5$. It is worth noting that the simplex bound (drawn for $A = 5$) is less than $1\,dB$ off from the two 32-point signal constellations shown in Figure 2. Another interesting observation is that the cutoff rate for the rectangular and the constant envelope signal constellations practically coincide, i.e., the signaling advantage, which, say 16-QAM has over 16-PSK on a AWGN-channel has disappeared.

Figure 3 shows the cutoff-rate for 8-PSK with the duty cycle ρ as a parameter. If the receiver is capable of extracting the jammer state side information – which it can achieve by measuring the ambient noise if N_J/N_0 is large – the jammer gains nothing by shortening its duty cycle for $E_S/N_J < 10\,dB$. Because the receiver knows the current jammer state it will simply blank out the jammed signals and have enough redundancy to recover the message. For higher values of E_S/N_J the jammer achieves some degradation of the communication link by shortening its duty cycle in accordance with (22). The achieved degradation is small however compared to uniform noise jamming ($\rho = 1$), if the receiver has information about the jammer state.

If the receiver is unable to extract side information from the received signal, i.e., the present jammer state s_τ is not known, it will have to use the ordinary Gaussian correlation metric

$$m(\mathbf{x}, \mathbf{y}) = \sum_{\tau=0}^{l} |\underline{y}_\tau - \underline{x}_\tau|^2, \tag{28}$$

and the Chernoff bound (8) becomes

$$C(\mathbf{x}, \mathbf{x}', \lambda) = \min_{\lambda} \underset{s}{E} \left[\prod_{\tau=1}^{l} e^{-\lambda(\underline{x}_\tau - \underline{x}'_\tau)^2 (1 - \lambda N_{s\tau})} \right]. \tag{29}$$

This expression cannot be minimized analytically. However the Chernoff factors are given by

$$\begin{aligned} C(x_\tau, x'_\tau, \lambda) &= \underset{s}{E} \left[e^{-\lambda(\underline{x}_\tau - \underline{x}'_\tau)^2 (1 - \lambda N_{s\tau})} \right] \\ &= \rho e^{-\lambda(\underline{x}_\tau - \underline{x}'_\tau)^2 (1 - \lambda(N_J/\rho + N_0))} + (1-\rho) e^{-\lambda(\underline{x}_\tau - \underline{x}'_\tau)^2 (1 - \lambda N_0)} \\ &= e^{-\lambda(\underline{x}_\tau - \underline{x}'_\tau)^2 (1 - \lambda N_0)} \left(\rho e^{\lambda^2 (\underline{x}_\tau - \underline{x}'_\tau)^2 N_J/\rho} + 1 - \rho \right). \end{aligned} \tag{30}$$

Substituting (30) into the formula for the cutoff-rate R_0 we obtain

$$R_0 = -\log_2 \min_{\lambda} \sum_{m=1}^{A} \sum_{p=1}^{A} p(\underline{a}_m) p(\underline{a}_p) C(\underline{a}_m, \underline{a}_p, \lambda)$$

$$= -\log_2 \min_{\lambda} \left(\sum_{m=1}^{A} \sum_{p=1}^{A} p(\underline{a}_m) p(\underline{a}_p) e^{-\lambda(x_r - x'_r)^2 (1-\lambda N_0)} \left(\rho e^{\lambda^2 (x_r - x'_r)^2 N_J/\rho} + 1 - \rho \right) \right). \tag{31}$$

This expression for R_0 is a function of the jammer duty cycle ρ.

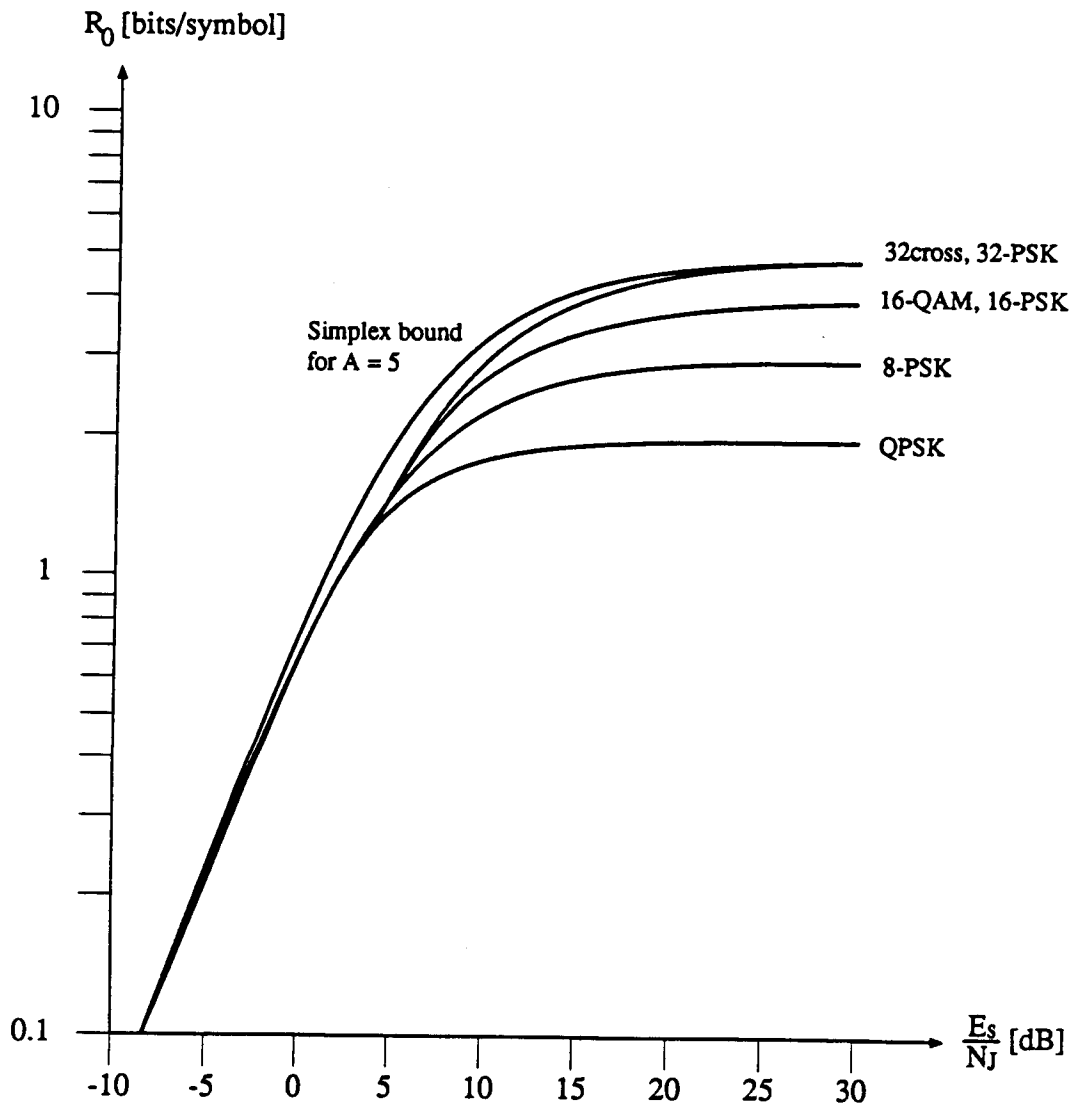


Figure 2: Cutoff-rate R_0 for some popular signal constellations and the simplex upper bound. ρ is chosen so as to minimize R_0.

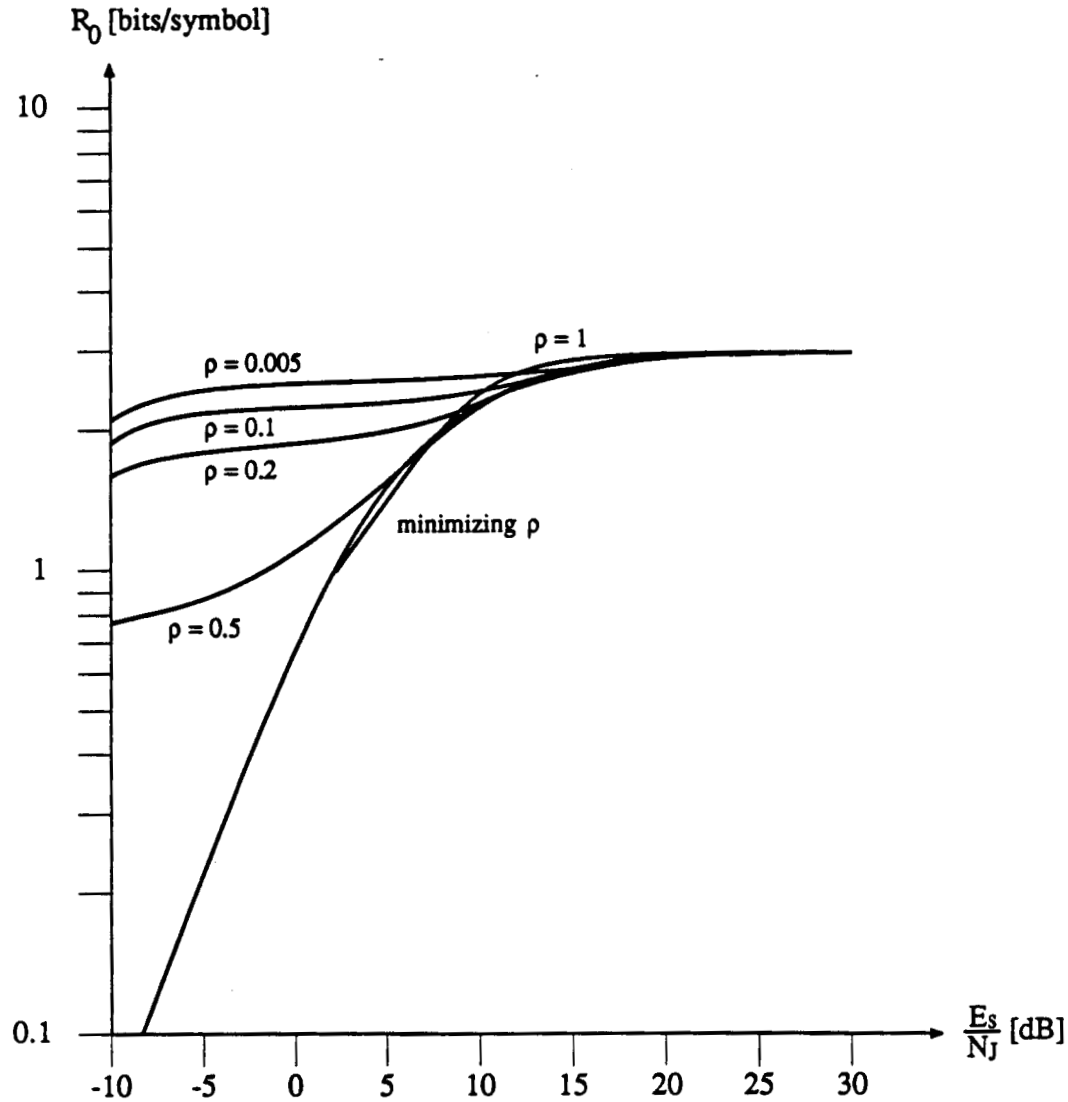


Figure 3: Cutoff-rate of 8-PSK for various values of the jammer duty cycle ρ.

Figure 4 shows R_0 without side information for a jamming power-to-noise ratio $N_J/N_0 = 100$. If the receiver has no jammer state side information, shortening the duty cycle ρ is detrimental at all values of E_S/N_J. In fact there is no lower limit for ρ. A similar behavior of the cutoff-rate has been observed for the binary channel in [Omura-Levitt, 1982]. In practice, however, the decoder would modify the Gaussian correlation metric by limiting the maximum values which $m(\underline{y}_r, \underline{x}_r)$ can assume. This clipping somewhat reduces the catastrophic effect of a very short jammer duty cycle ρ , see also [El-Wailly-Costello, 1975]. It is evident, nonetheless, that obtaining side information about the jammer state is crucial to successful communication over a jammed channel.

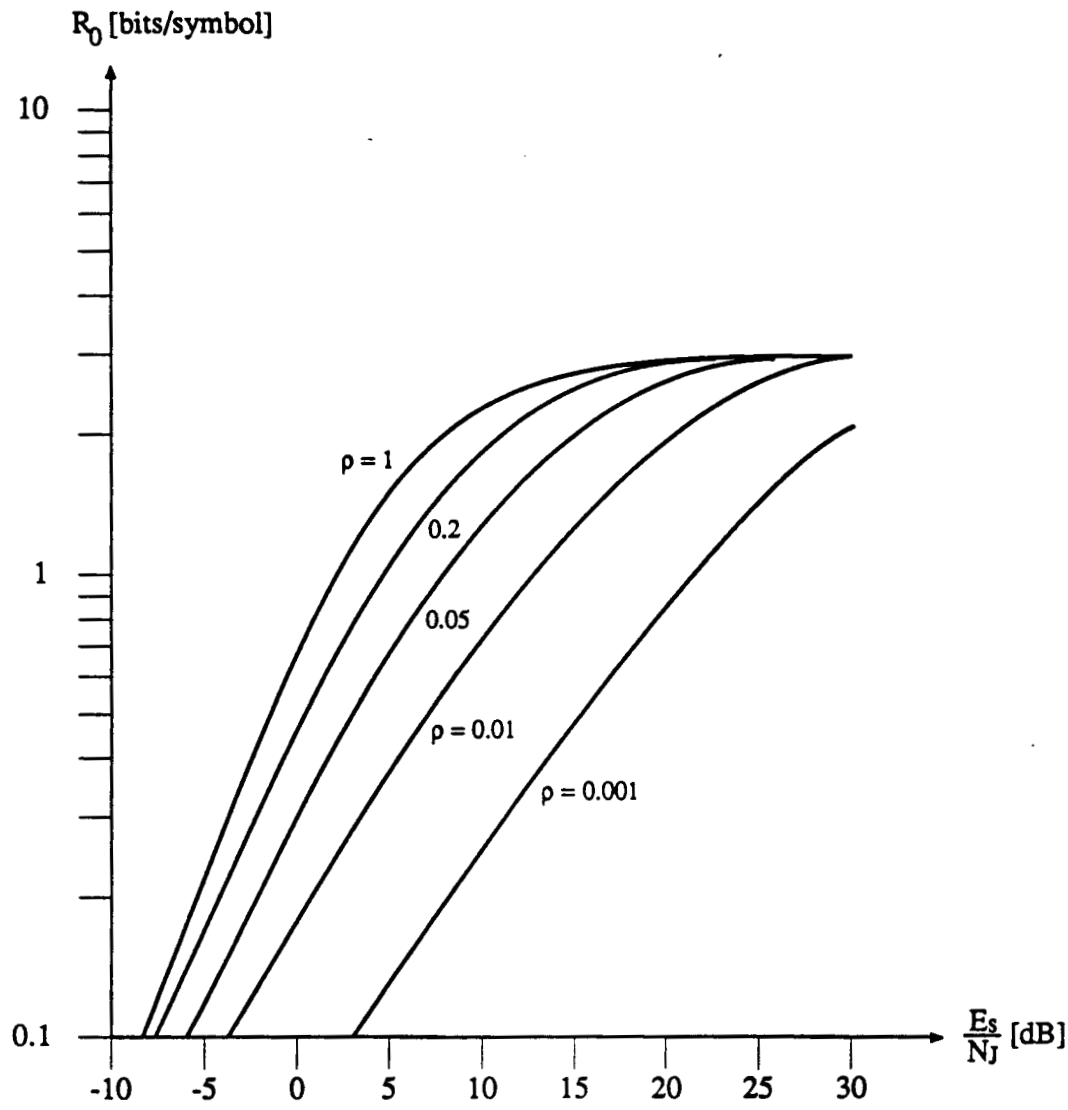


Figure 4: Cutoff-rate R_0 for 8-PSK for the jammed channel without jammer state side information.

4 Trellis Coded Modulation

The event error probability of TCM can be upper bounded by the transfer function bound [Schlegel, 1988]

$$P_e \leq \sum_t A_t P_t, \tag{32}$$

where the sum is over all code sequence pairs $\mathbf{x}, \mathbf{x}'$ whose two code word error probability is overbounded by the bound P_t, where $P_t = \prod_{\tau=1}^{t} C(x_\tau, x'_\tau)$ and A_t is the average number of code sequence pairs with Chernoff bound P_t.

Among the parallel transition error possibilities there will be a dominant term, P_C, given by

$$P_C = C(\underline{x}_\tau, \underline{x}'_\tau) = \frac{2(1-1/A)}{e^{\Delta^2(1-1/A)/2}} \frac{N_J}{E_S}, \tag{33}$$

where $\triangle^2$ is the smallest normalized Euclidean distance between signals on parallel transitions. From this it is evident that parallel transitions severely degrade code performance because P_C varies only inversely proportional with E_S/N_J.

For trellis error events, the dominant term, P_I, is given by

$$P_I = \prod_{\tau=1}^{l} C(x_\tau, x'_\tau) = \left(2(1-1/A)\frac{N_J}{E_S}\right)^{l'} \exp(-2(1-1/A)d(\mathbf{x},\mathbf{x}')), \tag{34}$$

where $d(\mathbf{x},\mathbf{x}') = \sum_{\tau=0}^{l} d_{\tau\tau'}$ is the normalized Euclidean distance of the signals $(\mathbf{x},\mathbf{x}')$ on the correct and incorrect path. In this case, the dominant code parameter is the effective length $l'_m = \min(l')$ of the code, where l' is the length of two code sequences less the number of time periods where they both are assigned identical signals. A trellis code designed for jammed channels must have a large effective length and a large Euclidean distance on the path pairs having $l' = l'_m$. Note that the minimum free Euclidean distance d_{free}, the dominant parameter for trellis codes on Gaussian channels, is of minor importance for trellis codes used on jammed channels. This is so because the path pair $(\mathbf{x},\mathbf{x}')$ that achieves d_{free} might have an effective length $l' > l'_m$. Table 1 shows the effective length and the Euclidean distances of Ungerboeck's 8-PSK codes.

i.d.	ν	$H(D)^0$	$H(D)^1$	$H(D)^2$	l'_m	$d(\mathbf{x},\mathbf{x}')$	d_{free}
u2	2	5	2	-	1	4	4
u4	4	23	04	16	3	6	5.17
u5	5	45	16	34	2	6	5.75
u6	6	103	030	066	3	8	6.34
u7	7	277	054	122	4	6.58	6.58
u8	8	435	072	130	3	8	7.52
u9	9	1007	164	260	3	8	7.52
u10	10	2003	164	770	4	10	7.52

Table 1: Ungerboeck's 8-PSK codes on the jammed channel. The connector polynomials $H(D)^j$ are defined in [Ungerboeck, 1982].

Table 2 shows a list of 8-PSK codes designed for the jammed channel. The code design method is discussed in [Schlegel, 1988].

i.d.	ν	$H(D)^0$	$H(D)^1$	$H(D)^2$	l'_m	$d(\mathbf{x},\mathbf{x}')$	d_{free}
j4	4	23	04	16	3	6	5.17
j5	5	43	14	36	3	8	5.17
j6	6	103	036	154	4	8	4.34
j7	7	223	076	314	4	8	4.93
j8	8	673	336	164	5	9.14	6.34
j9	9	1413	756	244	5	10.58	6.34
j10	10	3303	1676	504	5	14	6.34
j11	11	6403	3436	1264	6	11.17	7.52

Table 2: 8-PSK codes designed for the jammed channel.

Figures 5 - 7 show the transfer function bound (32) on the error event probability of some interesting codes on jammed channels. The bound was evaluated using the methods discussed in [Schlegel-Costello, 1988]. The codes are u5, u8, j5 and j8. Note that even though the codes j5 and j8 have a smaller minimum free Euclidean distance d_{free} than their Gaussian counterparts, their effective length and their minimum Euclidean distance $d(\mathbf{x},\mathbf{x}')$ for path pairs with effective length $l' = l'_m$ is larger, $d(\mathbf{x},\mathbf{x}') = 8$ for j5 versus $d(\mathbf{x},\mathbf{x}') = 6$ for u5 and $d(\mathbf{x},\mathbf{x}') = 9.14$ for j8 versus $d(\mathbf{x},\mathbf{x}') = 8$ for u8.

Figure 5 shows the event error probability of these four selected trellis codes on a most effectively jammed channel with jammer state side information, i.e., the jammer duty cycle ρ is given by (22). The special codes considerably outperform the Gaussian codes. At an error level of $P_e = 10^{-6}$ for example, j8 gains $5.5\,dB$ over u8. At $P_e = 10^{-4}$ j8 gains $2.8\,dB$ over u8 and j5 gains $4.5\,dB$ over u5. It becomes evident that for this jammed channel d_{free} is of secondary importance. u8 and j5 have the same effective length $l'_m = 3$ and $d(\mathbf{x}, \mathbf{x}') = 8$, so their asymptotic error curves have the same slope.

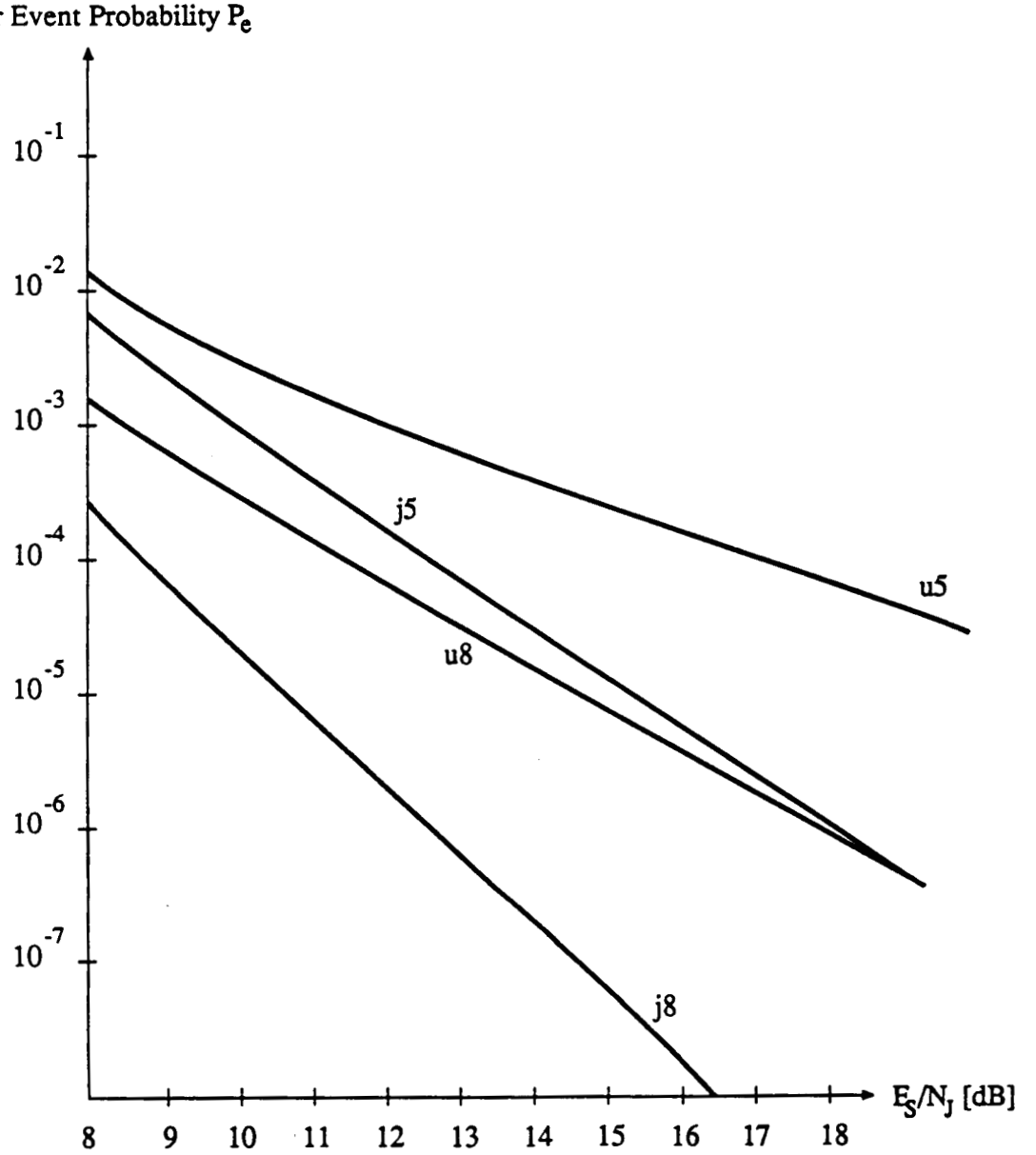


Figure 5: Performance of trellis codes on a most effectively jammed channel with channel state side information available at the receiver.

Figure 6 shows the event error probabilities for the four trellis codes on a jammed channel with duty cycle $\rho = 0.1$ and no jammer state side information. Note that the jamming noise power is only 10 times the power used by a uniform jammer. The error curves therefore show a steeper slope which is typical for Gaussian noise. It is interesting to note however that j8 and j5 still have a better error performance than u8 and u5, emphasizing the importance of a good effective length even in moderate impulse noise.

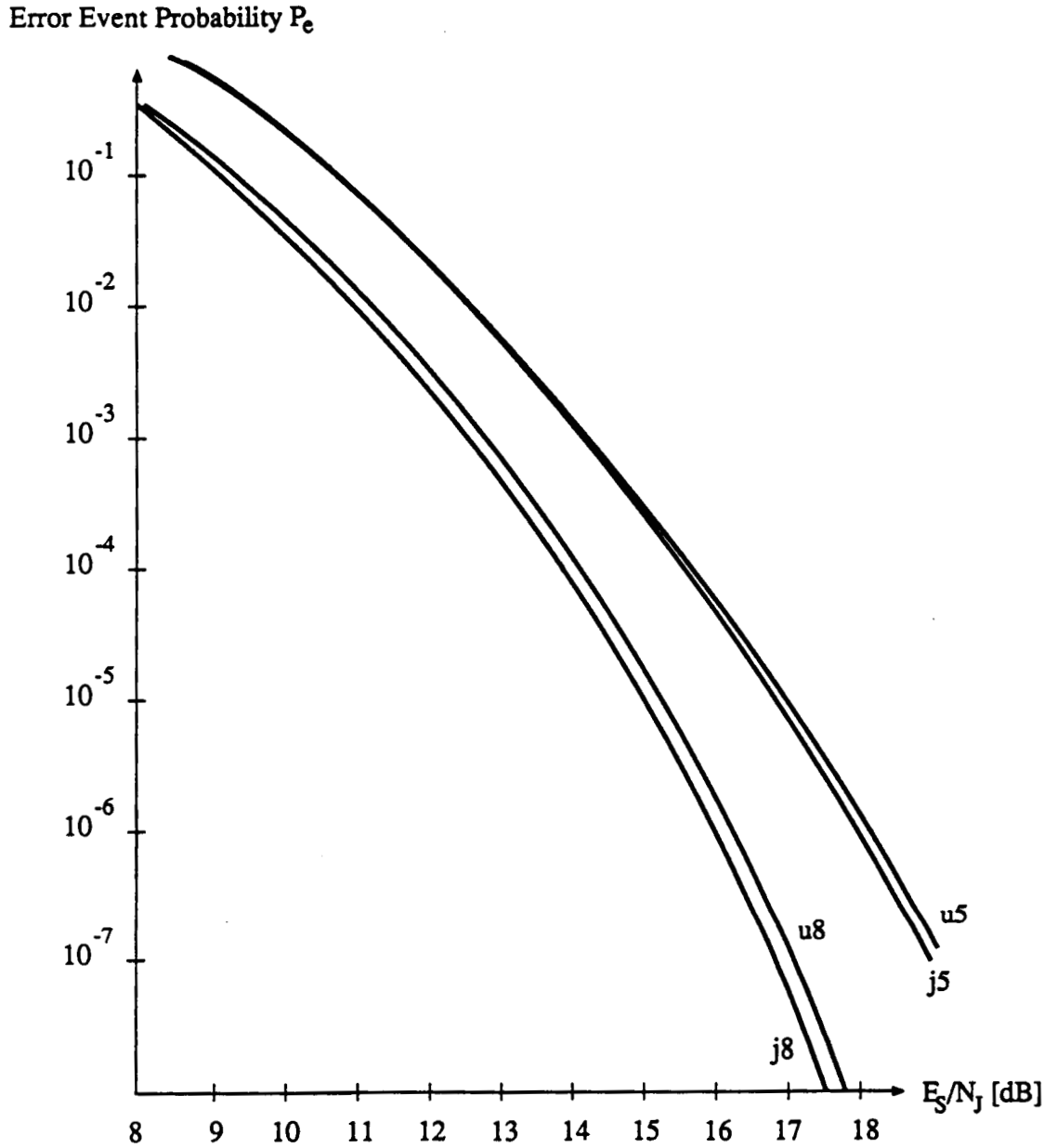


Figure 6: Performance of trellis codes on a jammed channel with jammer duty cycle $\rho = 0.1$ and without channel state side information.

If the jammer further shortens the duty cycle ρ, the performance of trellis codes without jammer state side information can be degraded almost arbitrarily as illustrated in Figure 7 for $\rho = 0.005$. The value of ρ characterizes severe impulse noise which not only flattens the error curves but also translates them along the E_S/N_J axis. Note that the E_S/N_J scale in Figure 7 is shifted by $10\,dB$ compared to the previous figures. The performance difference between j8 and u8 and j5 and u5 has also increased. It becomes evident in any case that side information and a good effective length l'_m of the code are essential for reliable communication on channels with impulse noise.

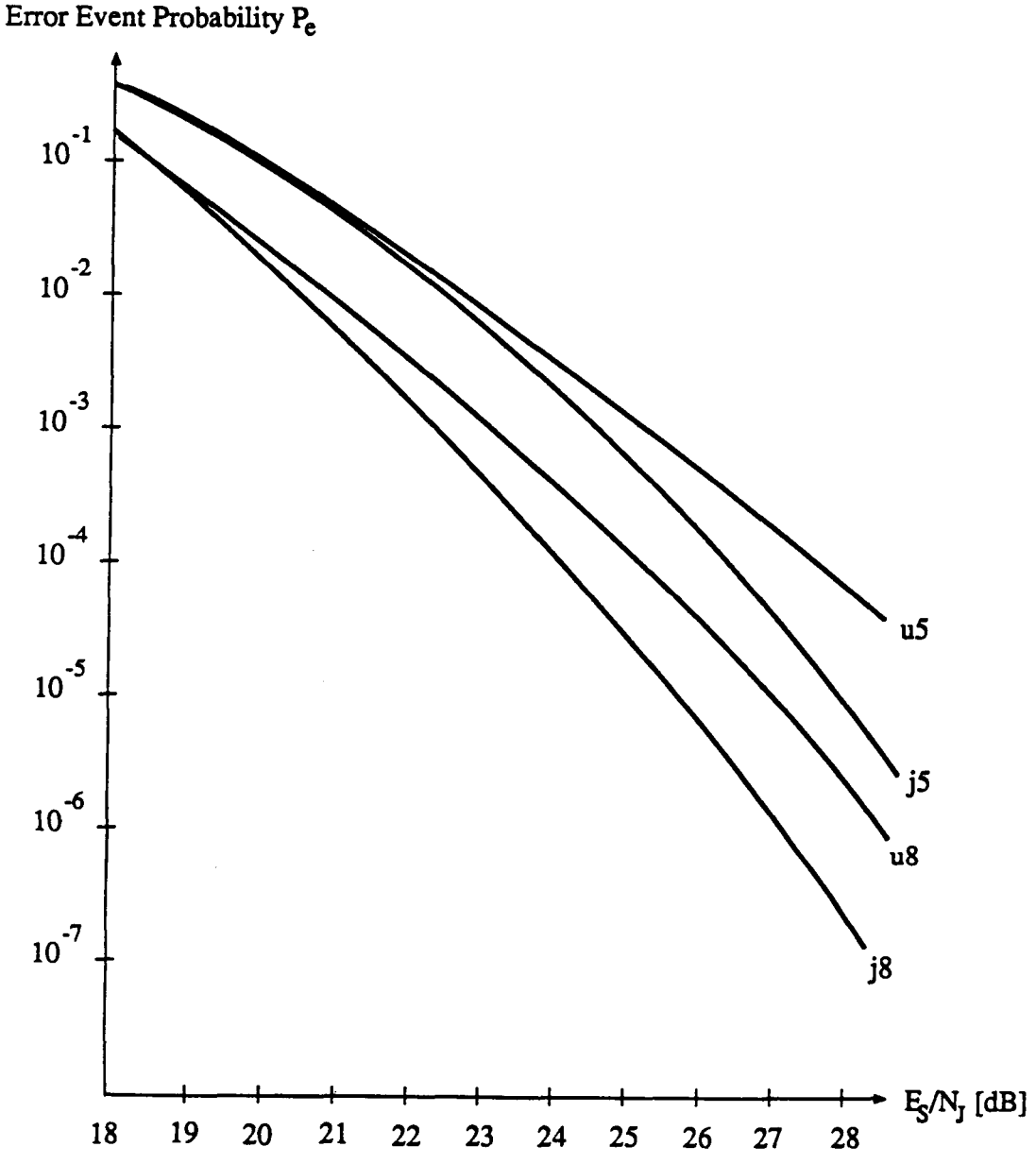


Figure 7: Performance of trellis codes on a jammed channel with jammer duty cycle $\rho = 0.005$ and without channel state side information.

5 Conclusions

We have analyzed the performance of TCM on impulse noise channels. For the worst case interferer, the intentional non-uniform jammer, the performance of a trellis code is dominated by its effective length. We showed through an error bound analysis that the effective lenght retains its importance even for moderate impulse noise. A number of codes were presented that perform better on impulse noise channels than their corresponding codes which were designed for AWGN channels.

References

[El-Wailly-Costello, 1975] F.F. El-Wailly and D. J. Costello, Jr., "Approximate Maximum Likelihood Decoding for a Coded System with Pulse Jamming," Private Communication.

[Gallager, 1968] R. G. Gallager, *Information Theory and Reliable Communication*, New York, Wiley, 1968.

[Massey, 1974] J. L. Massey, "Coding and Modulation in Digital Communications ", *Proc. Int. Zŕich Sem. Digital Commun.*, Zürich, Switzerland, March 1974, pp. E2(1)–E2(4).

[Omura-Levitt, 1982] J. K. Omura and B. K. Levitt, "Coded Error Probability Evaluation for Antijam Communication Systems," *IEEE Trans. Commun.*, Vol. COM-30, No.5, pp. 896–903, May, 1982.

[Schlegel, 1988] C. Schlegel, "Bandwidth Efficient Coding for Non-Gaussian Channels", Ph. D. Dissertation, University of Notre Dame, Notre Dame, Indiana, 1988.

[Schlegel-Costello, 1988] C. Schlegel and D. J. Costello, Jr., "Bandwidth Efficient Coding for Fading Channels," *IEEE Journal Selected Areas Commun.*, to appear 1989.

[Ungerboeck, 1982] G. Ungerboeck, "Channel Coding with Multilevel/Phase Signals," *IEEE Trans. Inform. Theory*, Vol. IT-28, No.1, pp. 55–67, January, 1982.

Appendix C

A Construction Algorithm for Trellis Codes

A CONSTRUCTION ALGORITHM FOR TRELLIS CODES

MARC ROUANNE, DANIEL J. COSTELLO, Jr.
Dept. of Elec. & Comp. Engr
Univ. of Notre Dame, Notre Dame, IN 46556

Abstract: A fast construction algorithm for good convolutional and trellis codes is presented. The technique can be used with any practical constraint length (up to about 30), any signal constellation, and any code rate. The construction does not require an exhaustive search, so that the complexity of the construction does not increase significantly with the constraint length. The codes constructed have slightly sub-optimum or optimum free distance.

Introduction: A trellis code, or trellis coded modulation (TCM), consists of a convolutional encoder followed by a mapper. Figure 1 shows a typical trellis code as originally designed by Ungerboeck [1][2]. An encoder state is characterized by the values of the past information bits stored in the shift registers of the convolutional encoder. The incoming information bits determine the transitions or branches from one state to the other. Subset selectors depend on the incoming bits and on ν past information bits only, where ν is the constraint length of the code. The mapper transforms the subset selector into a subset of signals and uncoded information bits select one particular signal from this subset. The free distance d^2_{free} is the minimum squared Euclidean distance between any two code words (sequences of signals).

There exist very few methods for constructing good convolutional or trellis codes [3][4]. Traditionally, short constraint length codes have been found by exhaustive searches. For larger constraint lengths, when the complexity of exhaustive searches becomes overwhelming, the searches are performed over restricted sets of codes. The algorithm presented in this paper searches over very small sets of codes (about 10 codes).

Most construction techniques and search algorithms can be used on expurgated sets of codes [5], where the expurgation guarantees a good distance growth from both ends of error events. Expurgated bounds on the free distance of trellis codes guarantee that such sets of codes contain codes with a large free distance [5]. This is a consequence of the observation that good convolutional or trellis codes must have good distance profiles from both ends of error events. Note that, like quick-look-in and complementary codes [3][4], Ungerboeck codes belong to expurgated sets of codes that achieve a maximum minimum distance between branches that reach and leave the same state [1].

Henceforth, we consider partition chains of the signal set which are composed of two-way partitions [5]. For simplicity, we suppose that the trellis has no uncoded bits (if the code has

This work was supported by NASA grant NAG5-557.

uncoded bits, then signals can be replaced with subsets of signals). Figure 2 shows a trellis code with maximum growth of the distance along the first q and last q branches of error events where $q = \lfloor n/k \rfloor$. The double line represents k parallel input bits and each memory cell represents k parallel memory cells also called a column of the encoder. Each set of k outputs selects k levels of the partition chain. Note that most of the best known trellis codes have the structure shown in Figure 2 [1] and [6]-[8].

If δ_j is the minimum distance between signals in the same subset at level j in the partition chain ($1 \le j \le n$), and if $\delta_0 \le \delta_1 \le \ldots \le \delta_n$, then the free distance of the codes shown in Figure 2 satisfies

$$d_{free}^2 \ge 2\delta_{n-k}^2 + 2\delta_{n-2k}^2 + \ldots + 2\delta_{n-qk}^2.$$

The codes are constructed from both ends simultaneously. The beginning of the code is constructed column by column, and each column is constructed from top to bottom. Similarly, the end of the code is constructed column by column, but each column is constructed from bottom to top [5]. If the code is large enough, q columns are constructed at both ends of the code. If the constraint length of the code is not large enough to accommodate $2q$ columns, then the columns are constructed in the following order: first the first and last columns, then the next $q - 1$ left columns, and finally the next $q - 1$ right columns. The construction stops when the desired constraint length is attained.

Only a few coefficients of the generator polynomials are chosen at each step (these coefficients correspond to a few memory cells of the encoder). The number of codes in the search increases linearly with the constraint length. Basically, the complexity of the algorithm is determined by the complexity of computing the free distance. Therefore, any class of codes whose free distance can be calculated can be constructed using this algorithm.

The algorithm is composed of two parts, an initialization phase and an iterative phase. The initialization of the algorithm guarantees a large distance along the first and last branches of error events. The codes have the structure shown in Figure 2. The iterative phase constructs rate k/n codes of constraint length $\nu + i$ from only one rate k/n code of constraint length ν, where $i \le k$. The iterative phase of the Algorithm is an exhaustive search over a very limited set of codes (2^n, the number of signals in the signal set).

The initialization of the algorithm consists in constructing the ends of the codes shown in Figure 2. The iterative part of the algorithm consists in extending the shortest registers of the code. For simplicity, the algorithm is described for rate $1/n$ trellis codes which are long enough to have the structure shown in Figure 2.

Initialization step:

Construct the rate $1/n$ trellis code with constraint length $\nu = 2q - 1 = 2n - 1$ as in Figure 2. Set j to q (the variable j indicates the position of the cell to be added to extend

the code).

Iterative steps:

(1): Given j, to construct a rate $1/n$ trellis code with constraint length $\nu + 1$ from a rate $1/n$ code with constraint length ν insert a memory cell at position j and let it have all possible connections to the n outputs of the binary encoder (2^n codes).

(2): If this memory cell is going to be the j^{th} of the next code go to (3) (good forward distance profile). If it is going to be the $(j + 1)^{th}$ memory cell of the next code, go to (4) (good backward distance profile).

(3): Among the codes with the best free distance in the set of 2^n codes, select the one which has the best distance profile along the first $j + 1$ branches and the best distance profile along the last $(\nu + 1) - j$ branches of error events. Increment j. Go to (1) unless the desired constraint length is reached.

(4): Among the codes with the best free distance in the set of 2^n codes, select the one which has the best distance profile along the first j branches and the best distance profile along the last $(\nu + 1) - j - 1$ branches of error events. The value of j is unchanged. Go to (1) unless the desired constraint length is reached.

Rate k/n trellis codes can be constructed by embedding k of this algorithm. Table 1 shows rate 1/2 codes. These codes are almost optimum for many constraint lengths and have free distances very similar to the free distances of quick-look-in codes and complementary codes even though the construction is much simpler. All these codes, except for very short constraint lengths have the structure shown in Figure 2. Table 2 gives rate 2/3 trellis coded 8-PSK schemes. These codes have good distance profiles from both ends of error events (although not optimum).

Recently, Porath and Aulin proposed a "fast algorithmic construction" to construct longer rate $k/(k + 1)$ codes by extending B shorter codes with an optimum distance profile (B is increased until the set of codes is large enough to contain a good code) [9]. This algorithmic construction is not fully deterministic unless B is large (the larger B is, the more exhaustive the search) and searches over many more codes than our algorithm. Modifying our algorithm to increase the size of the set of codes examined almost always gives optimum codes. However, we are more interested in finding fast algorithms to construct long codes than in finding optimum codes in a non-deterministic way.

Conclusion: The main criterion used to construct codes is the free distance, as opposed to other known algorithms which maximize the distance profile of codes. From a theoretical point of view, the construction technique shows that iterative construction algorithms can give codes with large free distance. It also shows that a compromise must be made between optimum distance profile and optimum free distance codes. From a practical point of view, it allows the fast construction of good codes for any TCM scheme. This can be the starting point for the construction of trellis codes optimum for specific applications.

References

[1] G. Ungerboeck, "Channel Coding with Multilevel Phase Signals", IEEE Trans. Inform. Theory, IT-28, pp. 55-67, Jan. 1982.

[2] G. D Forney, Jr., R. G. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Qureshi, "Efficient Modulation for Band-limited Channels", IEEE J. on Selected Areas in Communication, SAC-2, no. 5, pp. 632-647, Sept. 1984.

[3], J. L. Massey and D. J. Costello, Jr., "Nonsystematic Convolutional Codes for Sequential Decoding in Space Applications", IEEE Trans. Commun. Technol. , COM-19, pp. 806-813, October 1971.

[4] L. R. Bahl and F. Jelinek, "Rate 1/2 Convolutional Codes with Complementary Generators", IEEE Trans. Inform. Theory , vol. IT-17, No. 6, pp. 718-727, November 1971.

[5] M. Rouanne, "Distance bounds and construction algorithms for trellis codes", Ph.D Thesis, Notre Dame, 1988.

[6] L. F. Wei "Trellis-Coded Modulation with Multi-Dimensional Constellations", IEEE Trans. Inform. Theory, IT-33, pp. 483-501, July 1987.

[7] A. R. Calderbank and N. J. A. Sloane, "New Trellis Codes", IEEE Trans. Inform. Theory IT-33, pp. 177-195, March 1987.

[8] G. D. Forney, Jr., "Coset Codes I: Geometry and Classification", IEEE Trans. Inform. Theory, to appear.

[9] J. E. Porath and T. Aulin, "Fast Algorithmic Construction of Mostly Optimum Codes" Tech. Report no 5, Chalmers Univ. Techn., Goteborg, Sweden, February 1987.

[10] D. J. Costello, Jr., "A Construction Technique for Random-Error-Correcting Convolutional Codes", IEEE Trans. Inform. Theory, IT-15 , pp. 631-636, September 1969.

[11] K. J. Larsen, "Short Convolutional Codes with Maximum Free Distance for Rates 1/2, 1/3, and 1/4", IEEE Trans. Inform. Theory , IT-19, pp. 371-372, May 1973.

[12] G. Ungerboeck, "Trellis-Coded Modulation with Redundant Signal Sets. Part II: State of the Art", IEEE Comm. Mag. Vol. 25, pp. 12-21, February 1987.

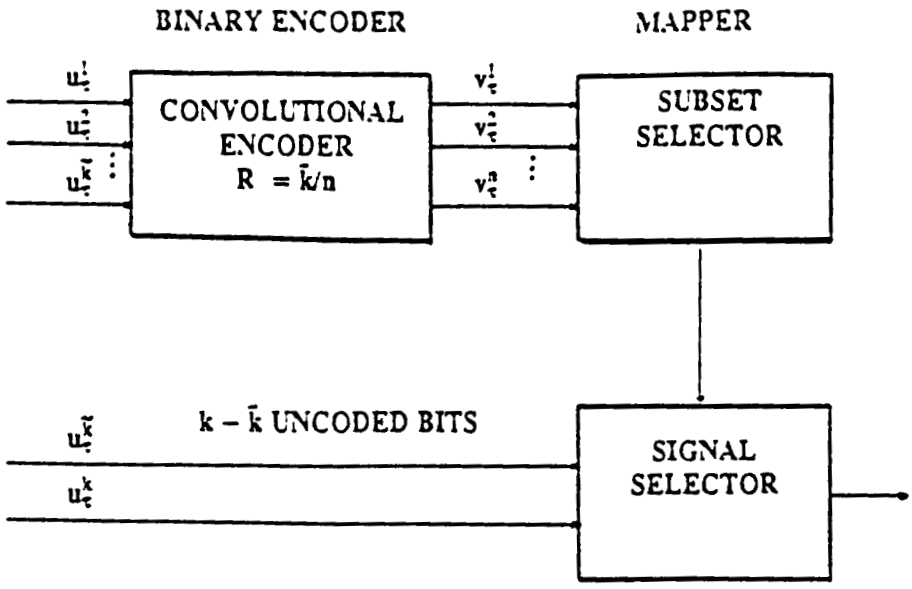


Figure 1: Schematic representation of a trellis code.

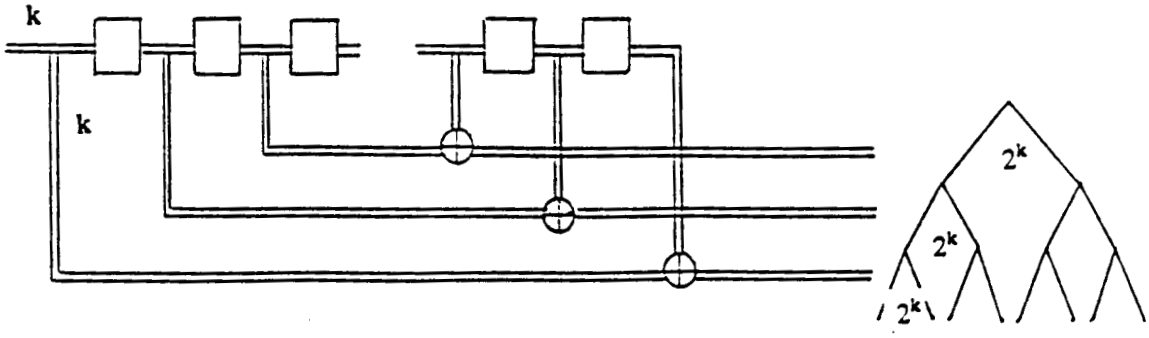


Figure 2: Trellis code with maximum growth of the distance at the ends of error events

Table 1: Construction of rate 1/2 convolutional codes.

ν	code	g_1	g_2	dir	d^1_{free}	C[2]	QLI[3]	BJ[4]	PA[5]	Max[6]
1	$f_2 f_1$	3	2	-	3	-	-	-	-	-
2	$f_2 f_1 f_2$	7	5	2f	5	4	5	5	5	5
3	$f_2 f_1 f_1 f_2$	17	11	2b	6	4	6	6	6	6
4	$f_2 f_1 \cdots f_1 f_2$	33	25	3f	7	5	7	7	7	7
5	$f_2 f_1 \cdots f_1 f_2$	63	55	3b	8	6	8	8	8	8
6	$f_2 f_1 \cdots f_1 f_2$	143	135	4f	9	6	9	9	10	10
7	$f_2 f_1 \cdots f_1 f_2$	313	265	4b	10	7	9	10	10	10
8	$f_2 f_1 \cdots f_1 f_2$	613	565	5f	10	7	10	11	11	12
9	$f_2 f_1 \cdots f_1 f_2$	1433	1345	5b	12	8	11	12	12	12
10	$f_2 f_1 \cdots f_1 f_2$	3073	2705	6f	12	8	12	13	13	14
11	$f_2 f_1 \cdots f_1 f_2$	6173	5605	7f	14	8	13	14	14	15
12	$f_2 f_1 \cdots f_1 f_2$	14373	13405	6b	14	9	14	15	14	16
13	$f_2 f_1 \cdots f_1 f_2$	30673	27105	7b	16	10	14	16	16	16
14	$f_2 f_1 \cdots f_1 f_2$	61473	56105	8f	16	10	15	17	16	18
15	$f_2 f_1 \cdots f_1 f_2$	143473	134505	8b	18	10	16	18	-	19
16	$f_2 f_1 \cdots f_1 f_2$	307073	271105	9f	18	10	16	18	-	20

1 Hamming free distance (search over 4 codes).
2 Construction technique [10] (search over 4 codes).
3 QLI codes [3] (search over $2^{\nu-1}$ codes).
4 Complementary codes [4] (search over $2^{\nu-1}$ codes).
5 Construction algorithm [9] (search over 32 codes).
6 Optimum codes [11] (exhaustive search).

Table 2: Construction of rate 2/3 trellis coded 8-PSK schemes.

ν	g_1	g_2	g_3	d^2_{free}	U1[1]	PA[9]	U2[12]
2	2 0	5 0	0 1	4	4	4	4
3	2 0	1 2	4 1	4.58	4.58	4.58	4.58
4	2 0	4 1	1 6	5.17	5.17	5.17	5.17
5	2 2	4 3	1 17	5.76	5.76	5.17	5.76
6	6 0	12 1	3 16	6.34	6.00	6.00	6.34
7	6 0	12 1	3 36	6.34	6.34	6.34	6.59
8	16 0	26 20	7 13	6.93	6.93	6.93	7.52
9	32 0	52 1	13 32	6.93	7.52	6.93	-
10	36 0	56 5	17 62	7.52	7.52	7.52	-
11	36 0	56 5	17 152	7.52	-	8.10	-
12	66 0	126 5	27 152	8.34	-	8.34	-

Appendix D

Construction of Good Distance Profile Trellis Codes for Sequential Decoding

CONSTRUCTION OF GOOD DISTANCE PROFILE TRELLIS CODES FOR SEQUENTIAL DECODING

SANKAR S.MALLADI
Dept. of Elect. Engr.
Univ. of Notre Dame
Notre Dame, IN 46556

DANIEL J.COSTELLO Jr.
Dept. of Elect. Engr.
Univ. of Notre Dame
Notre Dame, IN 46556

HENDRIK C.FERREIRA
Laboratory for Cybernetics
Rand Afrikaans Univ.
P.O.Box 524, Johannesburg
2000 South Africa

ABSTRACT

In this paper, we construct 2,4, and 8-dimensional trellis codes with rate k/k+1 systematic binary convolutional codes and multi level / phase signal sets which have good distance profiles using a nested step by step search algorithm. We then simulate the Fano algotithm to study the effects of the distance profile on the computational load of a sequential decoder. The results of few such decoding simulations are presented.

INTRODUCTION

In some practical applications where the available signal energy is relatively high, a Viterbi decoder at the receiving end of the communication system would be very inefficient. This is because, even for high Signal to Noise Ratio(SNR), it still performs 2^M computations per decoded information block, where M is the encoder memory. On the other hand, a sequential decoder, whose computational speed is adaptable to varying noise levels, would decode much faster on a high SNR channel.

This work was supported by NASA grant NAG5-557.

Another advantage of using a sequential decoder is that the decoding speed is independent of the encoder memory, whereas the number of computations performed by a Viterbi decoder increases exponentially as the encoder memory increases. This allows long constraint length codes to be used with a sequential decoder.

The performance of a sequential decoder depends on the Column Distance Function(CDF) of the code being used. Also, it is the initial part of the CDF, the Distance Profile, which affects the performance more than the latter part. These results for binary convolutional codes were first discovered by Chevillat and Costello[1].

DEFINITIONS

A rate k/k+1 systematic binary convolutional encoder of the form shown in figure[1(a)] can be described by a binary generator vector G(i), where

$$G(i) = [\ g^1(i),.....,g^k(i)\ ]^T, \qquad i = 0,1,.....,M,$$

and $g^h(i)$, $1 \leq h \leq k$, is equal to 0 if there is no connection to a mod-2 adder and equal to 1 if there is a connection and T stands for the transpose of the matrix. Figure[1(a)] shows the straight forward form of a rate2/3 systematic binary convolutional encoder and figure[1(b)] shows an alternate form the same encoder. Here, $G(0) = [0\ 0]^T$, $G(1) = [1\ 0]^T$, $G(2) = [0\ 0]^T$, $G(3) = [1\ 0]^T$, and $G(4) = [1\ 1]^T$. The encoder in figure[1(b)] can now be described by

$$G = \begin{bmatrix} g^{(1)} \\ \vdots \\ g^{(k)} \end{bmatrix} = [\, G(M) \ldots\ldots G(0) \,]$$

In figure[1(b)], $g^{(1)} = [\, 1\ 1\ 0\ 1\ 0 \,]$ and $g^{(2)} = [\, 1\ 0\ 0\ 0\ 0 \,]$

The CDF of a trellis code can be defined in the same way as for a binary convolutional code by replacing the Hamming weights by the Euclidean weights[2]. The CDF of order i, d_i , is defined as

$$d_i = \underset{\underline{x} \neq \underline{x}'_0}{\text{Min}} \left[\sum_{n=0}^{i} | a(\underline{z}_n) - a(\underline{z}'_n) |^2 \right]$$

$$\geq \underset{\{\underline{e}_0\} \neq \underline{0}}{\text{Min}} \left[\sum_{n=0}^{i} \Delta^2_{q(\underline{e}_n)} \right]$$

where $\{\underline{x}_n\}$ and $\{\underline{x}'_n\}$ are two distinct information sequences, $\{\underline{z}_n\}$ and $\{\underline{z}'_n\} = \{\underline{z}_n + \underline{e}_n\}$ are the corresponding code sequences , $\{\underline{e}_n\}$ is the error sequence, $\{a(\underline{z}_n)\}$ and $\{a(\underline{z}'_n)\}$ are the corresponding signal sequences, $\Delta^2_{q(\underline{e}_n)}$ are the Euclidean weights, and $q(\underline{e}_n)$ is the number of trailing zeros in $\{\underline{e}_n\}$. The distance profile of a trellis code can be defined as its CDF over only the first constraint length, i.e., $(d_0, d_1, d_2,..,d_M)$.

A trellis code is said to have a distance profile $(d_0,...d_M)$ superior to the distance profile $(d'_0,...,d'_M)$ of another code of the same memory order M if for some p, $0 \leq p \leq M$,

$$d_i = d'_i, \quad i = 0,1,2,....,p\text{-}1$$

$$> d'_i, \quad i = p.$$

Ungerboeck[3] constructed trellis codes for a Viterbi decoder. Later, Porath & Aulin [4] constructed mostly optimum trellis codes with either the same or a better d_{free}. In both cases the goal was to optimize the free distance, although in the latter case the algorithm also tries to acheive a good CDF. In our search for codes, the only goal was to optimize the distance profile. A combining algorithm to construct codes with good distance profile is also described in [4].

THE CODE SEARCH ALGORITHM

We employed a nested step by step search algorithm to construct codes with a good distance profile. Lin and Lyne[5] were the first to develop this approach which was also later used by Costello[6] and Hagenauer[7]. The code construction algorithm is outlined below.

Step 1:

Choose the initial k-bit connection vector which results in the maximum value of d_0. (It turns out that this vector is always G(0) = {0} for trellis codes.)

Step 2:

Store the information bits for the set of 2^k (one branch) paths along with their distances and the initial k-bit connection vector along with the distance d_0 and set i = 0, where i is the length of the shift registers.

Step 3:

Increment i and extend all the branches stored using each of the 2^k possible combinations of connection vectors G(i) and compute their distances d_i.

Step 4:

Choose the G(i) that results in the maximium growth of the distance profile. In case of a tie, the vector which resulted in the minimum number of paths with distance d_i is chosen.

Step 5:

Store the information bits for all paths of length i+1 branches along with their distances and the connection vector G(i) along with the distance d_i. If i = M, stop. Otherwise, repeat Step 3 until i = M, the encoder memory.

The same algorithm can be used to construct trellis codes using multi-dimensional modulation schemes. In this study, trellis codes using four and eight dimensional schemes with an 8-PSK signal constellation were constructed. The four-dimensional schemes were constructed using a rate 5/6 systematic binary convolutional encoder and the eight-dimensional schemes use a rate 11/12 encoder. Due to the fact that the algorithm requires an enormous computational time, encoders with uncoded bits were used. In each of the two cases, encoders with rates 2/3 and 3/4 were used. Figure [2] shows a rate 5/6, M = 4, systematic binary convolutional encoder with four uncoded bits for a four-dimensional 8-PSK scheme.

It should be noted that the number of paths whose distance must be computed increases exponentially with i. For this reason, a distance cutoff value is introduced and only those paths that do not exceed this cutoff value are stored for further extension. For short constraint length codes, a reasonably high cutoff value can be used to construct the codes along with their distance profiles. The cutoff value can be estimated from the distance profiles of the shorter constraint length codes using the following lemma.

Lemma:

For rate k/k+1 trellis coded modulation schemes,

$$d_{i+1} - d_i \leq \Delta_0, \quad i = 1,2,...,M-1.$$

Proof:

Consider the minimal form of the encoder shown in figure [1(b)]. Let d_j be the minimum distance at the j th stage.

Case 1. The content of the M th stage of the shift register is a 1. Then the signal selected would be from the subset $\{ S_1, S_3, ..., S_{2(k+1)-1} \}$, which would result in an increase of Δ_0 at the (j+1) th stage.

Case 2. The content of the M th stage of the shift register is a 0. Then, the signal selected would be from the subset $\{ S_0, S_2, ..., S_{2(k+1)-2} \}$, which would result in no increase at the (j+1) th stage.

The algorithm employs a straight forward binary encoder to construct the codes. The generators given in the tables[1-5] are for the form of the encoder shown in figure[1(b)].

The mapping of the k+1 bit binary code symbols z_n into the signal points $a(z_n)$ was done by using the set-partitioning method described in [3]. Trellis codes using rate 2/3 (8PSK) and 3/4 (16QAM) systematic binary convolutional codes were constructed. Codes using four and eight dimensional modulation schemes were also constructed.

We show in Tables[1-5] a list of 2, 4, and 8-dimensional trellis codes with rate 2/3 and rate 3/4 systematic binary convolutional codes using 8-PSK and 16 QAM signal constellations. A list of codes for the rate 3/4 (2 uncoded bits), 4-dimensional scheme is not included here since the algorithm found codes with $\mathbf{g}^{(3)} = \mathbf{0}$ and $\mathbf{g}^{(1)}$ and $\mathbf{g}^{(2)}$ the same as with 3 uncoded bits in table [3]. Figures [3 and 4] compare the distance profiles of two rate

2/3,constraint length 9and 10 codes respectively - one was constructed with the step by step algorithm presented here to optimize the distance profile and the other was constructed by Ungerboeck[3] to optimize the free distance. Figure[5] compares the distance profiles of two other codes, a good distance profile code constructed here and a mostly optimal code constructed with the combining algorithm[8], both of which are rate 2/3, constraint length 13 codes.

DECODING SIMULATIONS

To study how the distance profile affects the decoding speed or the number of computations required per frame by a sequential decoder, we simulated a Fano decoder and decoded noisy frames generated by using the rate 2/3 constraint length 9 and 10 codes and the rate 2/3, constraint length 13 code constructed with the step by step algorithm. We then used the rate 2/3 constraint length 9 and 10 codes constructed by Ungerboeck(UG) and the mostly optimal(MO) code of rate 2/3, and constraint length 13.

THE SIMULATION PROCEDURE

Step 1.

Generate a 128-bit long random sequence of 1's and 0's called a frame.

Step 2.

Encode the bits and map the binary encoder outputs into the 8-PSK signal points.

Step 3.

Generate random 2-dimensional Gaussian noise vectors and add them to the signal points.

Step 4.

Map the received signals into 8-PSK signal points with a hard decision demodulator based on the decision regions shown in figure [6].

STEP 5.

Decode the noisy frames with a Fano decoder.

Each simulation involved generating a thousand random sequences of 1's and 0's, encoding, and mapping the code words into an 8-PSK signal set. Each sequence was a frame of 128 symbols. A Gaussian channel was simulated to transmit these signal points. Hard decisions were made by the Fano decoder based on the optimum decision regions shown in figure[6]. Each forward look by the decoder was counted as one computation.The metric used was not an optimum metric, but was good enough to compare the computational distributions of the different codes. The bit metric used for decoding binary convolutional codes was used as the symbol metric for trellis codes:

$$M[r(\underline{z}_n)|a(\underline{z}_n)] = \log_2 [\, P(r(\underline{z}_n)|a(\underline{z}_n)) \,/\, P(r(\underline{z}_n))] - R,$$

where, $r(\underline{z}_n)$ is a received symbol, $a(\underline{z}_n)$ is a transmitted symbol, and R is the code rate.

The results of these simulations at different SNRs are shown in figures [7-12]. C is the number of computations required per frame. The figures show the probability that C is greater than or equal to a number N plotted against N. The results obtained generally agree with the results previously found using binary convolutional codes[1]. The code with the better distance profile performed fewer computations on the average.

CONCLUSIONS

The results of the decoding simulations were satisfactory and generally agreed with the results previously found for binary convolutional codes.The codes with a better distance profile performed fewer computations on the average. Also, the difference between the performance of the codes with different distance profiles seems to be greater than the difference for binary convolutional codes.

It can be concluded that when choosing a trellis code for a sequential decoder, it is important to choose codes with rapidly increasing distance profiles. Also, since the computational performance of the sequential decoder does not depend on the constraint length of the code, long constraint length codes which guarantee a large d_{free} can be used. This makes the free distance a secondary criterion when designing trellis codes for sequential decoding.

REFERENCES

1. P.R.Chevillat and D.J.Costello,Jr., "Distance and Computation in Sequential Decoding", IEEE Trans. Commun., COM-24, pp.440-447, April 1976.

2. G.Ungerboeck, "Trellis-Coded Modulation with Redundant Signal Sets. Part II: State of the Art", IEEE Commun Mag., vol.25, No.2,pp.12-21, Feb. 1987.

3. G.Ungerboeck, "Channel coding with Multi-level/Phase Signals", IEEE Trans. Inform. Theory, vol.IT-28, pp.55-67, Jan. 1982.

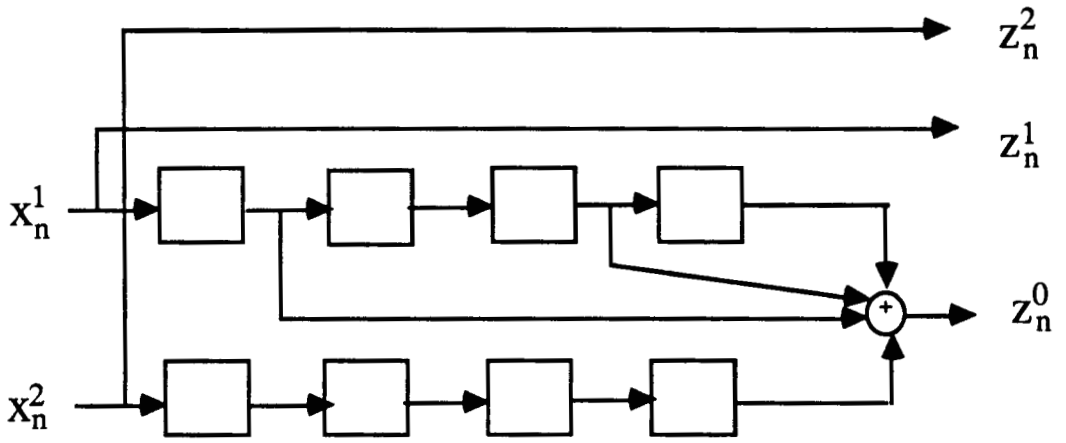
4. J.E.Porath and T.Aulin, "Fast Algorithmic Construction of Mostly Optimal Trellis Codes", Tech. Rep. No.5, Div. of Inform. Theory, School of ECE, Chalmers Univ. of Tech., Goteborg, Sweden, Feb. 1987.

5. S.Lin and H.Lyne, "Some results on Binary Convolutional Code Generators", IEEE Trans. Inform. Theory, vol.IT-13, pp.134-139, Jan. 1967.

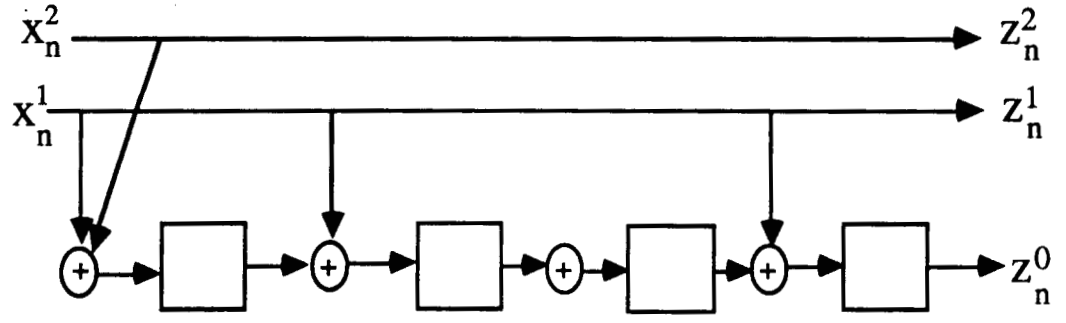
6. D.J.Costello,Jr., "A Construction Technique for Random Error Correcting Convolutional Codes", IEEE Trans. Inform. Theory, vol.IT-15, pp.631-636, Sept. 1969.

7. J.Hagenauer, "High Rate Convolutional Codes with Good Distance Profiles", IEEE Trans. Inform. Theory, vol.IT-23, pp.615-618, Sept. 1977.

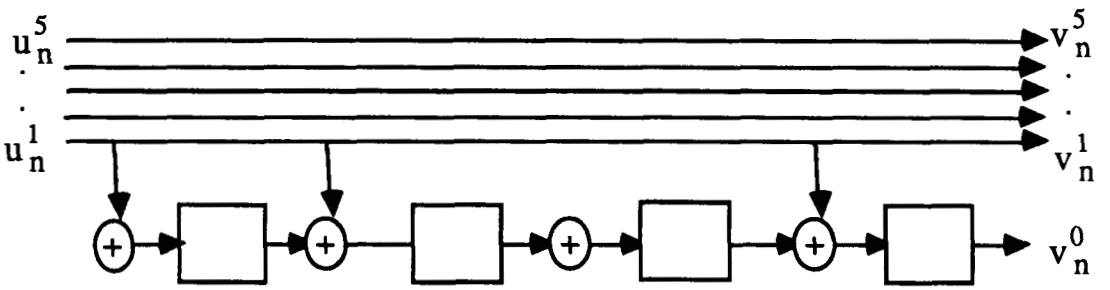
8. G.J.Pottie and D.P.Taylor, "A Comparision of Reduced Complexity Decoding Algorithms for Trellis Codes", CRL Rep. No.190, Faculty of Engr., McMaster Univ., Hamilton, Ontario, Canada, March 1988.



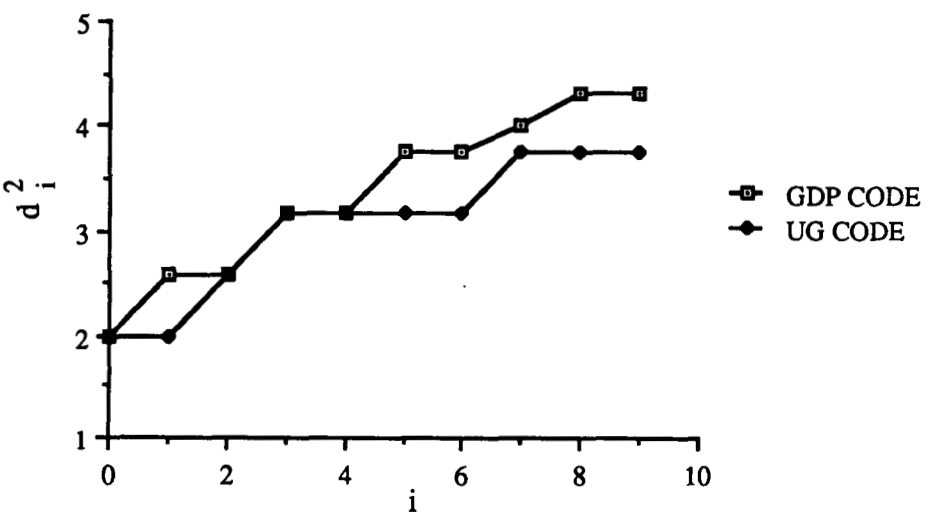
Fig[1(a)]. A straight forward realization of a rate 2/3 constraint length 4 convolutional code.



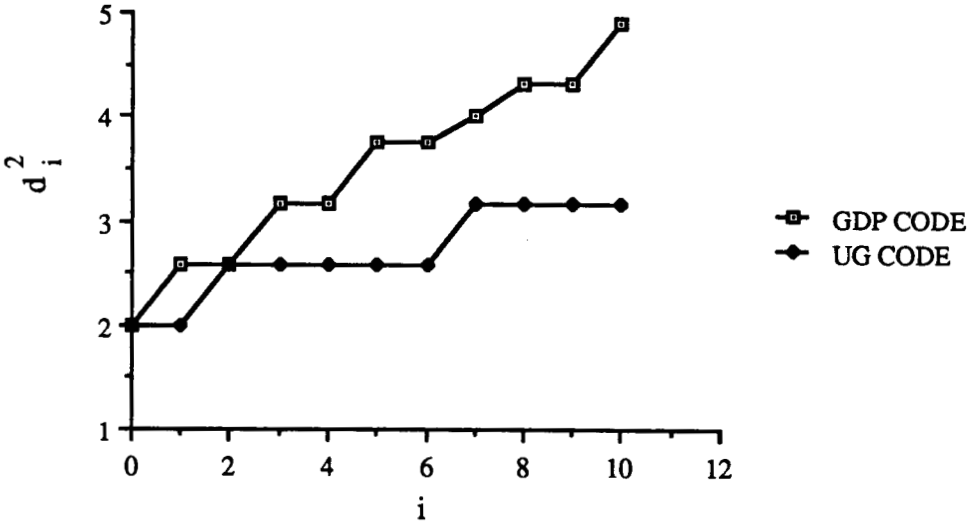
Fig[1(b)]. An alternate form of the encoder in figure[1(a)].



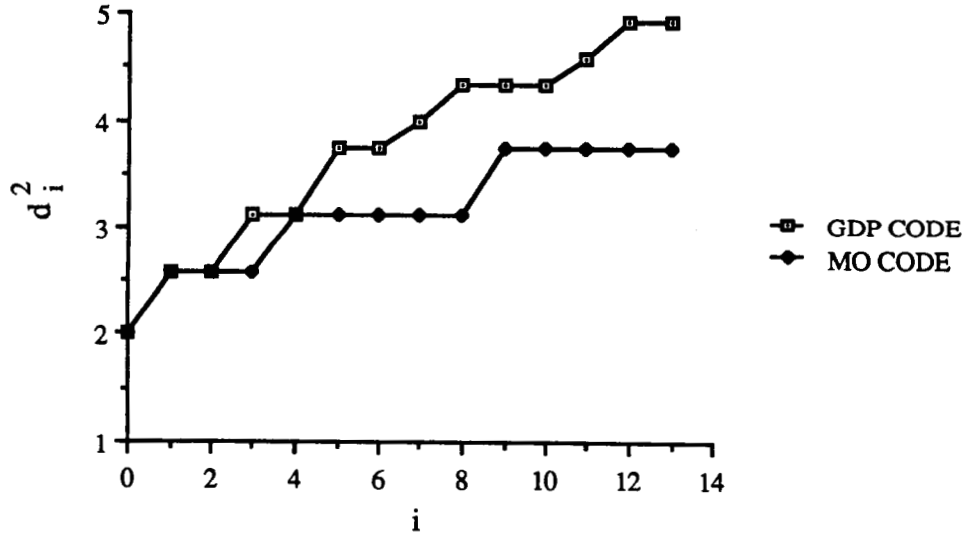
Fig[2]. A Systematic Rate 5/6, M = 4 Encoder for a 4-D TCM Scheme.



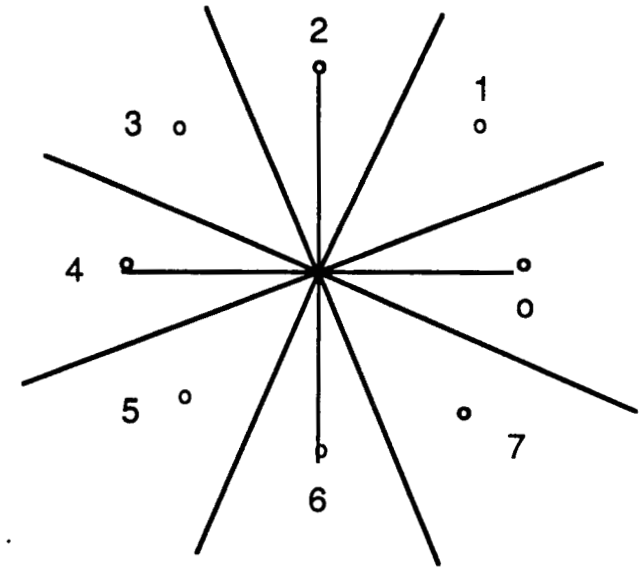
Fig[3]. Distance profiles of two rate 2/3, constraint length 9 trellis codes.



Fig[4]. Distance profiles of two rate 2/3, constraint length 10 trellis codes.



Fig[5]. Distance profiles of two rate 2/3, constraint length 13 trellis codes.



Fig[6]. 8-PSK signal set with the decision regions.

M	$g^{(1)}$	$g^{(2)}$	d_M^2
3	12	0	3.17
4	32	20	3.17
5	52	0	3.74
6	152	100	3.74
7	252	0	4.00
8	652	400	4.32
9	652	1400	4.58
10	652	3400	4.90
11	652	7400	4.90
12	652	17400	5.16
13	20652	37400	5.48
14	60652	77400	5.48
15	60652	177400	5.48
16	260652	177400	5.48
17	660652	577400	5.48
18	1660652	577400	5.48
19	3660652	577400	5.74
20	7660652	577400	5.74
21	3660652	10577400	6.02
22	23660652	10577400	6.02
23	63660652	10577400	6.02
24	163660652	110577400	6.33
25	163660652	210577400	6.33

Table[1]. A list of rate 2/3 trellis codes with 8-PSK signal set.(All generators are in octal.)

M	$g^{(1)}$	$g^{(2)}$	$g^{(3)}$	d_M^2
3	16	04	04	1.2
4	16	24	24	1.6
5	56	64	64	1.6
6	156	164	164	1.6
7	256	064	264	2.0
8	656	464	364	2.0
9	1656	0464	1364	2.0
10	3656	2464	1364	2.0
11	7656	6464	5364	2.4
12	17656	16464	15364	2.4
13	37656	16464	15364	2.4
14	37656	56464	15364	2.4
15	137656	056464	015364	2.4

Table[2]. A list of rate 3/4 trellis codes with 16-QAM signal set.(All generators are in octal.)

M	$g^{(1)}$	$g^{(2)}$	d_M^2
3	12	00	2.0
4	32	20	2.3
5	72	20	2.3
6	032	120	2.3
7	232	320	2.9
8	632	320	2.9
9	1632	0320	2.9
10	0632	2320	3.2
11	4632	4320	3.5
12	14632	04320	3.5
13	24632	04320	3.5
14	64632	04320	3.5
15	164632	004320	3.8
16	364632	204320	4.1
17	764632	204320	4.1
18	1764632	0204320	4.1
19	3764632	0204320	4.1
20	5764632	4204320	4.7
21	15764632	14204320	4.7
22	35764632	14204320	4.7
23	55764632	14204320	4.7

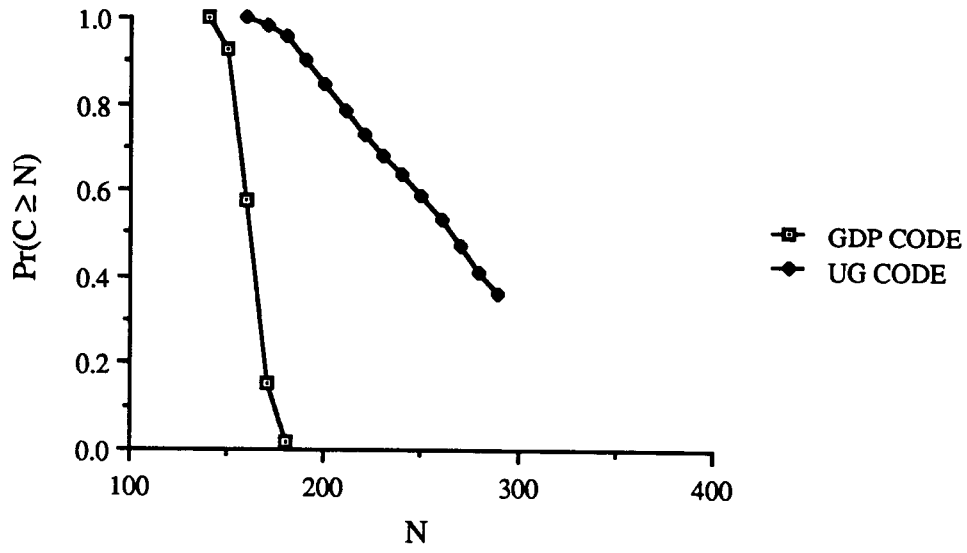
TABLE[3]. A LIST OF GOOD DISTANCE PROFILE 4-DIMENSIONAL TRELLIS CODES WITH AN 8-PSK SIGNAL SET AND A RATE 2/3 SYSTEMATIC BINARY CONVOLUTIONAL ENCODER (3 UNCODED BITS).

M	$g^{(1)}$	$g^{(2)}$	d_M^2
3	16	04	1.8
4	26	04	1.8
5	66	44	2.3
6	166	044	2.3
7	266	044	2.3
8	666	044	2.3
9	1266	1044	2.9
10	3266	1044	2.9
11	5266	5044	2.9
12	15266	05044	2.9
13	15266	25044	3.5
14	45266	25044	3.5
15	105266	125044	3.5
16	305266	125044	3.5
17	505266	125044	3.5

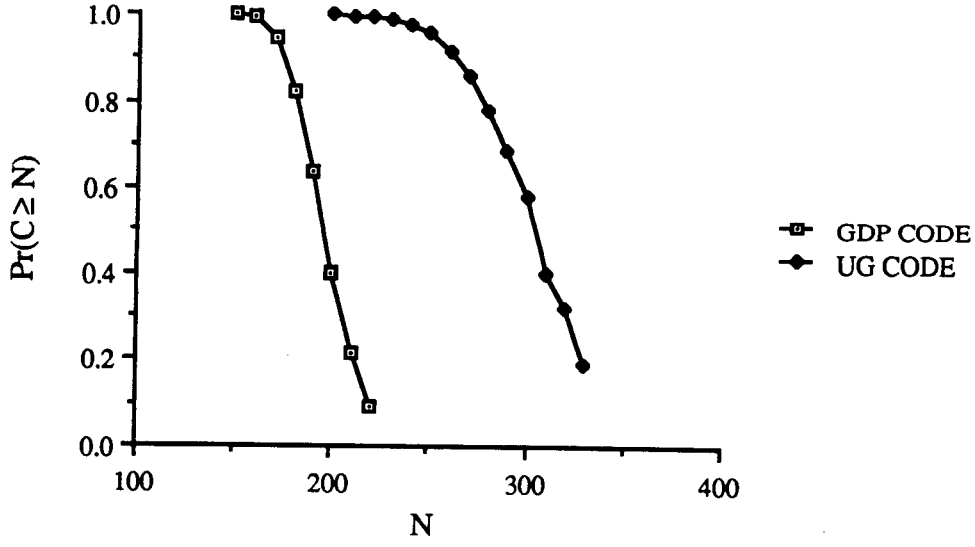
TABLE[4]. A LIST OF GOOD DISTANCE PROFILE 8-DIMENSIONAL TRELLIS CODES WITH AN 8-PSK SIGNAL SET AND A RATE 2/3 SYSTEMATIC BINARY CONVOLUTIONAL ENCODER(9 UNCODED BITS).

M	$g^{(1)}$	$g^{(2)}$	$g^{(3)}$	d_M^2
3	16	04	00	1.8
4	26	04	00	1.8
5	66	44	00	2.0
6	166	144	100	2.3
7	166	344	300	2.3
8	566	744	300	2.3
9	0566	1744	1300	2.3

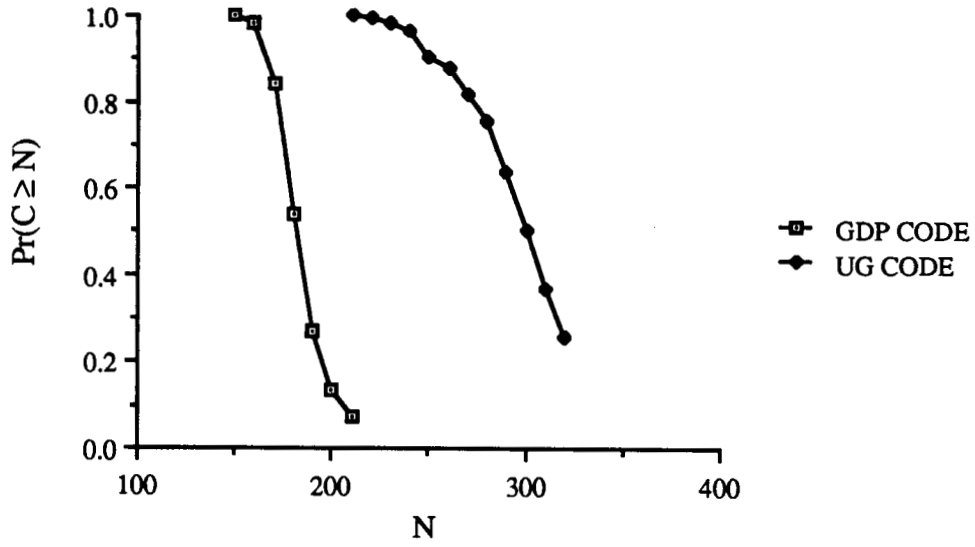
TABLE[5]. A LIST OF GOOD DISTANCE PROFILE 8-DIMENSIONAL TRELLIS CODES WITH AN 8-PSK SIGNAL SET AND A RATE 3/4 SYSTEMATIC BINARY CONVOLUTIONAL ENCODER(8 UNCODED BITS).



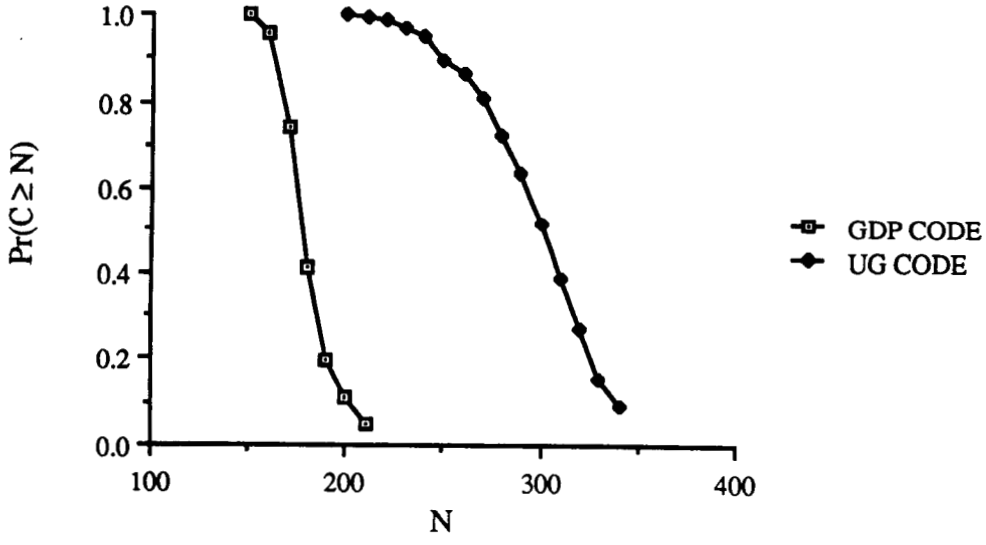
Fig[7]. Computational Distribution for two rate 2/3, constraint length 9 trellis codes at 11 dB.



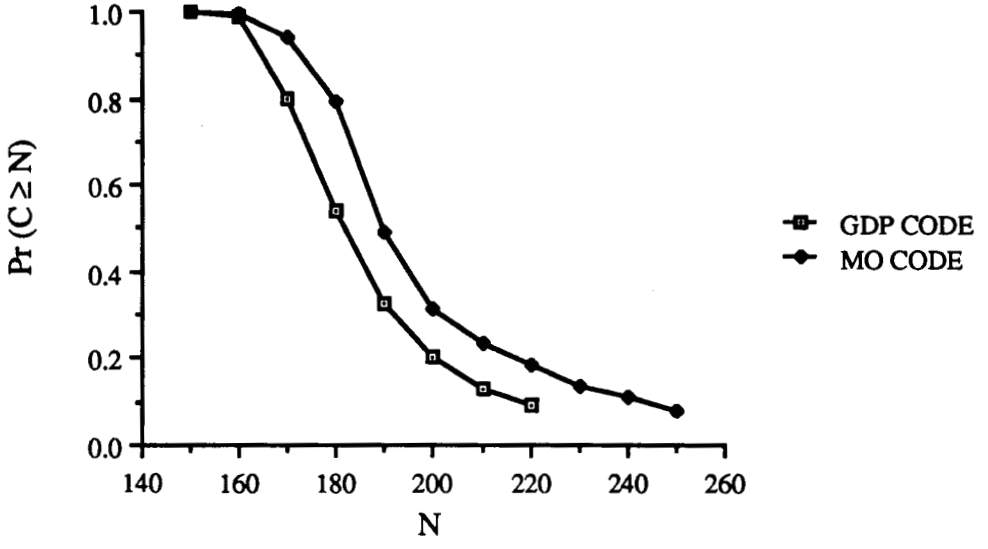
Fig[8]. Computational Distribution for two rate 2/3, constraint length 9 trellis codes at 8 dB.



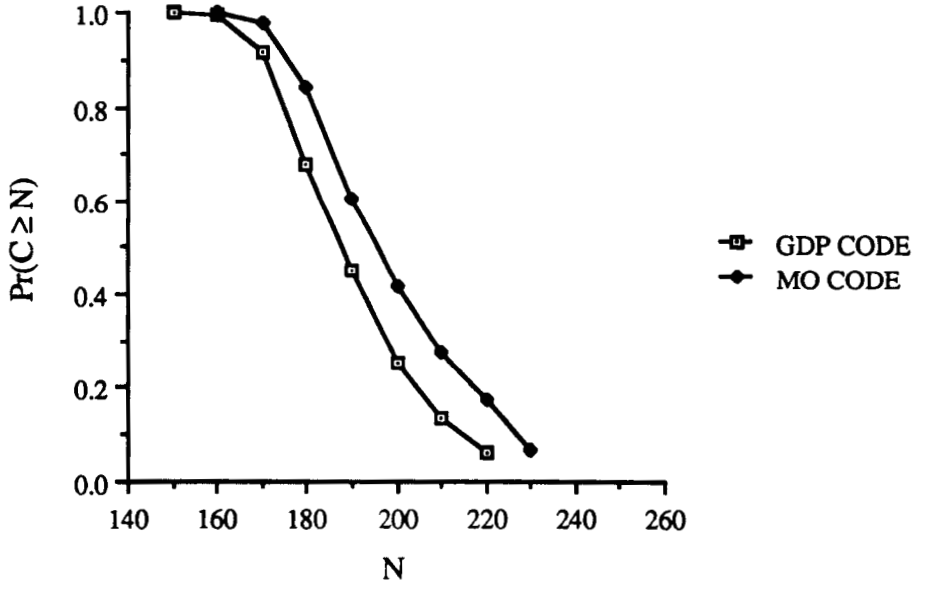
Fig[9]. Computational Distribution for two rate 2/3, constraint length 10 trellis codes at 8 dB.



Fig[10]. Computational Distribution for two rate 2/3, constraint length 10 trellis codes at 9 dB.



Fig[11]. Computational Distribution for two rate 2/3, constraint length 13 trellis codes at 10 dB.



Fig[12]. Computational Distribution for two rate 2/3, constraint length 13 trellis codes at 8 dB.

Appendix E

A Comparison of the Random Coding Bounds on the Free Euclidean Distance and the Event Error Probability of Trellis Coded Modulation Schemes

A COMPARISON OF THE RANDOM CODING BOUNDS ON THE FREE EUCLIDEAN DISTANCE AND THE EVENT ERROR PROBABILITY OF TRELLIS CODED MODULATION SCHEMES [1]

Christian Schlegel, Marc Rouanne and Daniel J. Costello, Jr.

Department of Electrical and Computer Engineering
University of Notre Dame
Notre Dame, Indiana 46556

Abstract: A comparison of an expurgated upper bound on the event error probability of trellis coded modulation schemes is compared to a lower bound on the achievable minimum free Euclidean distance. It is shown that the bounds are equivalent asymptotically and a geometric construction method for the expurgated error exponent is given. Several constellations are optimized with respect to the bounds.

1 Introduction

In recent years bandwidth efficient Trellis Coded Modulation (TCM) introduced by Ungerboeck [1] has become increasingly popular and much analysis has been devoted to the performance of these coding schemes on AWGN channels. It is well known that for large signal to noise ratios (SNR's) the minimum free Euclidean distance d_{free} is the dominant parameter in code performance. Much research has gone into the search for codes with large d_{free}. While most of this work focused on finding good trellis codes with a given signal constellation, the constellation itself is also a parameter in the system design. There have been a few attempts to design better codes by using non-standard signal constellations, usually asymmetric MPSK signal sets [2]. While these codes showed slight performance improvements, no general rule on how to choose the constellation is known.

2 Error Bounds from the Free Euclidean Distance

A rate $R = k/n$ trellis code is generated by a binary convolutional encoder followed by a mapper as shown in Figure 1. The convolutional encoder is a finite state automaton, which, at each time interval j, accepts k binary input bits, makes a transition from its state S_j at time j to one of 2^k possible successor states S_{j+1} and outputs one of 2^n binary n-tuples. In the minimal realization [3] the encoder consists of k shift registers with lengths $m_1, \ldots, m_k$. Let the memory order m be $\min(m_i)$ and assume henceforth that $m_i = m$ for all i. The mapper then translates each binary n-tuple into one of $A = 2^n$ channel signals from a (possibly multidimensional) signal set $\mathcal{A} = \{a_1, a_2, \ldots, a_A\}$. The signal sets discussed in this paper are *4-ary pulse amplitude modulation* (4-PAM), 8-PAM, *8-ary phase shift keyed modulation* (8-PSK), and *16 quadrature amplitude modulation* (16-QAM). These signal sets are shown in Figure 2. A rate $R = k/n$ trellis code transmits k bits/channel signal.

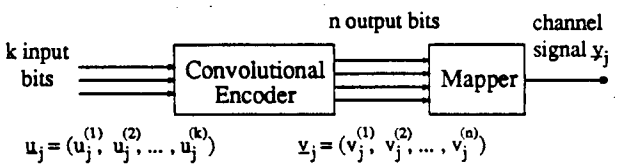


Figure 1: Convolutional Encoder and Mapper for a TCM-scheme.

At the receiver, the signal sequences are decoded probabilistically using the Viterbi algorithm. The Viterbi algorithm is a maximum likelihood sequence estimator that finds the code sequence which most closely corresponds to the sequence of received signals. Any decoding error in a trellis code is characterized by a set of incorrect branches which the decoder chooses over the correct branches. With each incorrect path we may associate a sequence of incorrect trellis states $\hat{S}_j$, while the sequence of correct states is S_j. An error event of length l can be described by l state pairs, $(S_0, \hat{S}_0), \ldots, (S_l, \hat{S}_l)$, with $S_0 = \hat{S}_0$,

[1]This work was supported by NASA Grant NAG5-557

$S_l = \hat{S}_l$, and $S_j \neq \hat{S}_j$ for $0 < j < l$, i.e., the incorrect path must not touch the correct path. Associated with these paths are two symbol sequences $\mathbf{y} = [\underline{y}_0, \underline{y}_1, \ldots, \underline{y}_{l-1}]$ and $\hat{\mathbf{y}} = [\underline{\hat{y}}_0, \underline{\hat{y}}_1, \ldots, \underline{\hat{y}}_{l-1}]$, where $\underline{y}_j$ and $\underline{\hat{y}}_j$ are signals from the signal set $\mathcal{A}$. The probability of an error event is a function of the cumulative Euclidean distance between the code sequences associated with the correct and the incorrect path. We may therefore write

$$\Pr[(S_0, \ldots, S_l) \rightarrow (\hat{S}_0, \ldots, \hat{S}_l)] = Q\left(\sqrt{\frac{||\mathbf{y} - \hat{\mathbf{y}}||^2}{2N_0}}\right), \tag{1}$$

where $Q(\gamma) = \int_\gamma^\infty 1/\sqrt{2\pi} e^{-\gamma^2/2} d\gamma$ is the complementary error function, N_0 is the one sided noise power spectral density, and $||\mathbf{y} - \hat{\mathbf{y}}||^2 = \sum_{j=0}^{l-1} |\underline{y}_j - \underline{\hat{y}}_j|^2$ is the norm of the difference of the two code sequences $\mathbf{y}$ and $\hat{\mathbf{y}}$.

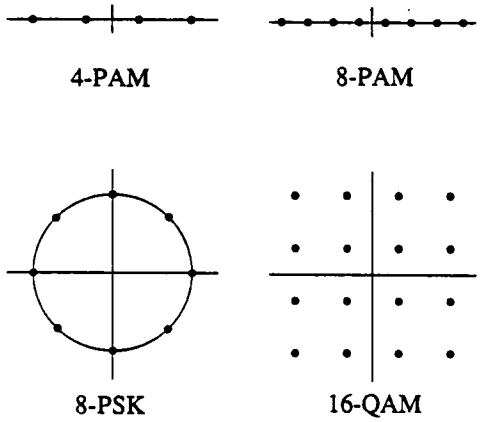


Figure 2: Constellations discussed in this paper.

In [4] it is shown that the event error probability P_e can be lower bounded by

$$P_e \geq \frac{1}{l} Q\left(d_{free}\sqrt{\frac{E_s}{2N_0}}\right), \tag{2}$$

where $d_{free} = \min_{\mathbf{y}, \hat{\mathbf{y}}} \sqrt{||\mathbf{y} - \hat{\mathbf{y}}||^2 / E_s}$ is the normalized minimum free Euclidean distance of the code and l is the length of the path that achieves d_{free}. E_s is the average signal energy $E_s = \sum_{i=1}^{A} p(\underline{a}_i)|\underline{a}_i|^2$. Using the well known approximation of the Q-function [6, page 83]

$$\frac{1}{\sqrt{2\pi}\gamma}(1 - \frac{1}{\gamma^2}) e^{-\frac{\gamma^2}{2}} < Q(\gamma) < \frac{1}{\sqrt{2\pi}\gamma} e^{-\frac{\gamma^2}{2}}, \tag{3}$$

we obtain for (2)

$$\begin{aligned} P_e &\geq \frac{1}{\sqrt{2\pi l}} \exp\left(-d_{free}^2 \frac{E_s/N_0}{4} - \ln\left(d_{free}\sqrt{\frac{E_s/N_0}{2}}\right)\right. \\ &\qquad \left. + \ln\left(1 - \frac{2}{d_{free}^2 E_s/N_0}\right)\right) \\ &= \frac{1}{\sqrt{2\pi l}} \exp\left(-d_{free}^2 \frac{E_s/N_0}{4}(1 + o(E_s/N_0))\right) \end{aligned} \tag{4}$$

where $o(E_s/N_0) \propto \ln(E_s/N_0)/(E_s/N_0)$ is a quantity that goes to 0 as $E_s/N_0 \rightarrow \infty$.

From the transfer function bound for trellis codes we know that the asymptotic value of the event error probability P_e for large E_s/N_0 is approximated by

$$P_e \approx n_{free} Q\left(d_{free}\sqrt{\frac{E_s/N_0}{2}}\right), \tag{5}$$

where n_{free} is the path multiplicity of the code, i.e., the number of distinct error events with the minimum free Euclidean distance d_{free}. The free distance d_{free} thus gives us a good measure of code performance at high values of E_s/N_0.

3 Random Coding Error Bounds

In [4], it is shown that there exists an expurgated upper bound on the event error probability of the best trellis code given by

$$P_e \leq \left(\frac{(2^k - 1)e^{-k\ln 2}}{e^{E(s) - k\ln 2} - 1}\right)^{\frac{1}{s}} e^{-m\frac{E(s)}{s}}, \quad E(s) > k\ln 2, \tag{6}$$

where $0 < s \leq 1$ and

$$E(s) = -\ln \sum_{i=1}^{A}\sum_{j=1}^{A} q(\underline{a}_i) q(\underline{a}_j) e^{-s\frac{|\underline{a}_i - \underline{a}_j|^2}{4} E_s/N_0}, \tag{7}$$

Let us further define the expurgated error exponent E_{ex} as

$$E_{ex} \triangleq \max_{0 < s \leq 1} \left(\frac{E(s)}{s}\right). \tag{8}$$

For large memory lengths m, E_{ex} is the exponent that determines code performance.

We thus have an upper bound (6) and a lower bound (4) on the event error probability. Since these two bounds must not contradict each other,

$$\frac{1}{\sqrt{2\pi l}}\exp\left(-d_{free}^2\frac{E_s/N_0}{4}(1+o(E_s/N_0))\right)\leq$$

$$\left((2^k-1)e^{-k\ln 2}\right)^{\frac{1}{s}}e^{-\frac{m}{s}E(s)-\frac{1}{s}\ln\left(e^{E(s)-k\ln 2}-1\right)}, \quad (9)$$

where d_{free} is now the minimum free Euclidean distance of the best trellis code. We may take the natural logarithm of both sides and obtain

$$\ln(\sqrt{2\pi}l)-d_{free}^2\frac{E_s/N_0}{4}(1+o(E_s/N_0))\leq$$

$$-\frac{m}{s}E(s)-\frac{1}{s}\ln\left(e^{E(s)-k\ln 2}-1\right)+\frac{1}{s}\ln\left(1-e^{-k\ln 2}\right)$$

$$d_{free}^2(1+o(E_s/N_0))\geq\frac{4m}{sE_s/N_0}E(s)$$

$$+\frac{4}{sE_s/N_0}\ln\left(e^{E(s)-k\ln 2}-1\right)+\frac{4}{sE_s/N_0}\ln\left(1-e^{-k\ln 2}\right)$$

$$+\frac{4\ln(\sqrt{2\pi}l)}{E_s/N_0}. \quad (10)$$

For simplicity, let us denote $\frac{sE_s/N_0}{4}$ by α. Rewriting (10) above we obtain

$$d_{free}^2(1+o(E_s/N_0))\geq\frac{mE(\alpha)}{\alpha}+\frac{\Theta(E(\alpha))}{\alpha}$$

$$+\frac{4\ln(\ln(\sqrt{2\pi}l)}{E_s/N_0} \quad (11)$$

where

$$\Theta(E(\alpha)) = \ln\left(e^{E(\alpha)-k\ln 2}-1\right)-\ln\left(1-e^{-k\ln 2}\right)$$

$$E(\alpha) = -\ln\sum_{i=1}^{A}\sum_{j=1}^{A}q(\underline{a}_i)q(\underline{a}_j)e^{-\alpha|\underline{a}_i-\underline{a}_j|^2}, \quad (12)$$

$$0 \leq \alpha\leq\frac{E_s/N_0}{4}.$$

In order to obtain a lower bound on the minimum free Euclidean distance d_{free}, we let $E_s/N_0\rightarrow\infty$ in (11) and obtain the same bound as Rouanne [5], i.e.,

$$d_{free}^2\geq\max_{0\leq\alpha}\left(\frac{mE(\alpha)}{\alpha}+\frac{\Theta(E(\alpha))}{\alpha}\right); E(\alpha)>k\ln 2. \quad (13)$$

4 Asymptotic Error Estimates

For large m, the contribution of the term $\Theta(E(\alpha))$ will become negligible in (13) and we may therefore ignore it. Maximizing the minimum free Euclidean distance in (13) is then the same as maximizing the expurgated error exponent E_{ex}, given in (8), with the only difference being that the range of $\alpha(s)$ is restricted in (6). If the maximizing $\alpha_{max}\leq(E_s/N_0)/4$, then the resulting constellation that maximizes the bound on the free distance will be the same that maximized the expurgated error exponent, and

$$d_{free}^2\gtrapprox m\frac{E(\alpha_{max})}{\alpha_{max}} \quad (14)$$

as well as

$$P_e\leq\eta e^{-m\frac{E(\alpha_{max})}{\alpha_{max}}\frac{E_s/N_0}{4}}, \quad 0<\alpha_{max}\leq\frac{E_s/N_0}{4}, \quad (15)$$

where

$$\eta=\left(\frac{(2^k-1)e^{-k\ln 2}}{e^{E(\alpha_{max})-k\ln 2}-1}\right)^{\frac{E_s/N_0}{4\alpha_{max}}}. \quad (16)$$

Note that

$$\frac{E(\alpha_{max})}{\alpha_{max}}\frac{E_s/N_0}{4}=E_{ex}. \quad (17)$$

If $\alpha_{max}=(E_s/N_0)/4$, then the maximizing parameter in (8) will be $s=1$. The error bound (6) then becomes

$$P_e\leq\frac{(2^k-1)e^{-k\ln 2}}{e^{R_0-k\ln 2}-1}e^{-mR_0}, \quad R_0>k\ln 2, \quad (18)$$

where

$$R_0=-\ln\sum_{i=1}^{A}\sum_{j=1}^{A}q(\underline{a}_i)q(\underline{a}_j)e^{-\frac{|\underline{a}_i-\underline{a}_j|^2}{4}E_s/N_0} \quad (19)$$

is the cutoff-rate of the constellation in nats/signal. Computer evaluations of (8), or equivalently of (13) with $m\rightarrow\infty$, show that $E(s)=E(\alpha)\rightarrow k\ln 2$ for $\alpha\rightarrow\alpha_{max}$. This observation holds true for all constellations we have analyzed.

Figure 3 shows the expurgated error exponent E_{ex} for 4-PAM as a function of E_s/N_0. Since α_{max} is independent of E_s/N_0 (17) shows that E_{ex} is a linear function of E_s/N_0. The slope of this function, $E(\alpha_{max})/(4\alpha_{max})$, is a function of k because α_{max} depends on k. As E_s/N_0 approaches $4\alpha_{max}$ from above, $E_{ex}\rightarrow R_0$. $4\alpha_{max}$ is the smallest E_s/N_0 for which

there exists an upper bound of the form (6). The exponent of the upper bound at this point is R_0. The higher the available energy, i.e., the larger E_s/N_0 for a particular k, the larger the expurgated error exponent will be. For a code with a larger value of k, the expurgated exponent will grow more slowly with E_s/N_0 and a larger minimum E_s/N_0 is required to guarantee the exponential bound. R_0 retains its importance as a cutoff-rate, since it is not only the expurgated exponent at the cutoff-point $E_s/N_0 = 4\alpha_{max}$, but also the maximum rate $k = R_0/\ln 2$ for which (6) exists. Thus the expurgated error exponent as well as the asymptotic part of the free distance bound (13) can be easily *constructed* from a graph of the cutoff-rate R_0 in the following manner. The construction is illustrated in Figure 3 for rate $k = 1.99$ bits/symbol for 4-PAM signalling.

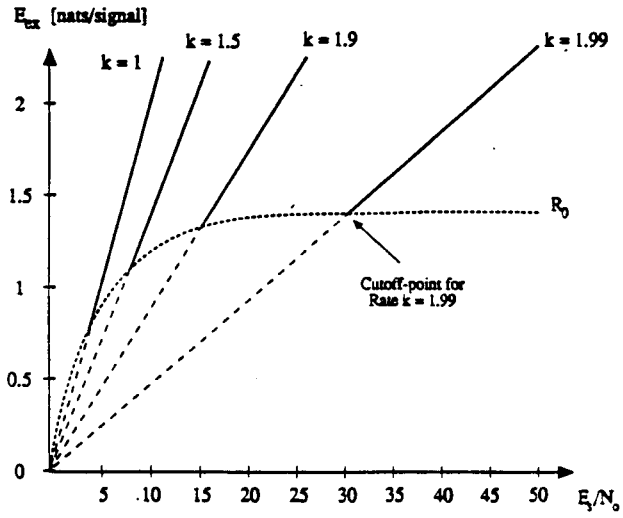


Figure 3: Cutoff-rate R_0 and expurgated error exponent of 4-PAM.

(1) Choose the value of k (which determines the rate either in nats/signal or bits/signal). The cutoff-point is the intersection of a line parallel to the E_s/N_0-axis at a distance k (in nats/signal) with the cutoff-rate curve. The x-value of the cutoff-point is $4\alpha_{max}$.

(2) Connect the origin of the graph with the cutoff-point in a line g.

(3) The expurgated exponent for any $E_s/N_0 > 4\alpha_{max}$ is the y-value of g at that E_s/N_0-value.

(4) The asymptotic bound on d^2_{free} from equation (14) is $4m\times$ the slope of g.

5 Constellation Optimization

An important parameter in (13) is the constellation itself, that is, the coordinates of the signal points in the set $\mathcal{A}$. We have already observed that for $m \to \infty$ and compatible values of k, the maximization of the minimum free Euclidean distance and the expurgated error exponent will yield the same signal constellation. The cutoff-rate R_0 of rectangular constellations (16-QAM etc.) is higher than R_0 of their constant envelope counterparts (16-PSK, etc.) in the moderate E_s/N_0-range. It should be obvious from the exponent construction that constellations with a better cutoff-rate at a given E_s/N_0 will also yield a better expurgated error exponent as well as a better bound on d_{free}.

Next we have optimized the constellations introduced in Figure 2 using a program implementing the Newton algorithm to find the global minimum of a function of n variables. Figure 4 shows the results for the 4-PAM constellation and the 8-PAM constellation for a number of rates. It is interesting to note that for small rates k both constellations converge towards binary signalling (BPSK).

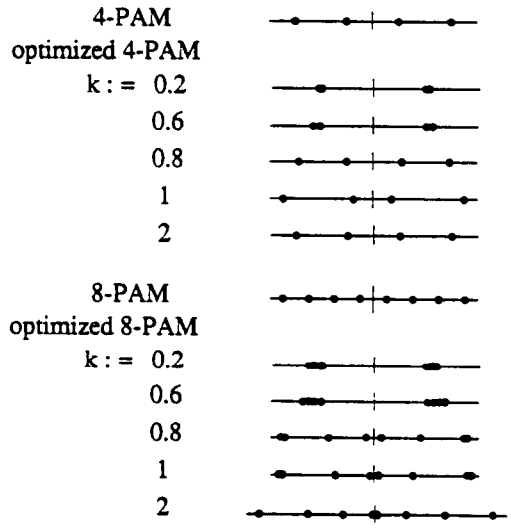


Figure 4: Optimized 4-PAM and 8-PAM constellations for different rates k.

Figure 5 shows the optimization for 8-PSK. Contrary to the PAM signal sets, the maximizing constellation does not vary with E_s/N_0. The angles are slightly shifted towards asymmetric 8-PSK.

Figure 6 shows an optimized 16-QAM for rate $k = 3$. As the two dimensions have been maximized under the same constraints, the constellation is the cartesian product of two 4-PAM constellations.

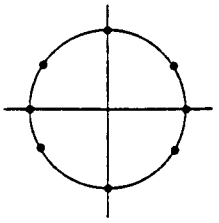


Figure 5: Optimized 8-PSK signal set.

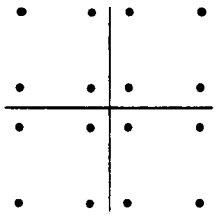


Figure 6: Example of an optimized 2-dimensional signal set for rate $k = 3$.

6 Conclusions

We have presented a link between the expurgated upper bound on the event error probability of a trellis code and a bound on its maximum achievable minimum free Euclidean distance d_{free}. We have shown that the expurgated error exponent and the asymptotic bound on d_{free} can be constructed from the cutoff-rate curve of a constellation. We have presented the optimization values of the error exponent for a number of popular constellations. It is apparent that for low rates, small constellations are preferable.

REFERENCES

[1] G. Ungerboeck, "Channel Coding with Multilevel/Phase Signals," *IEEE Trans. Inform. Theory*, Vol. IT-28, No.1, pp. 55–67, January, 1982.

[2] D. Divsalar and M. K. Simon, "Combined Trellis Coding with Asymmetric Modulations", *IEEE Trans. Commun.*, Vol. COM-35, No .2, pp. 130–141, February 1987.

[3] G. D. Forney, Jr., "Convolutional Codes I: Algebraic Structure," *IEEE Trans. Inform. Theory*, Vol. IT-16, No. 6, pp. 720–738, November, 1970.

[4] C. Schlegel, "Bandwidth Efficient Coding for Non-Uniform Channels", Ph.D Dissertation, University of Notre Dame.

[5] M. Rouanne and D. J. Costello, Jr., "A Lower Bound on the Minimum Euclidean Distance of Trellis Coded Modulation", *IEEE Trans. Inform. Theory*, to appear, 1988.

[6] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, New York, Wiley, 1965.